Vol. XXVII. Part I. February, 1957

The
British Journal
of
Educational Psychology

EDITED BY

P. E. VERNON

AIDED IN THE SELECTION OF PAPERS BY

SIR CYRIL BURT
R. A. C. OLIVER
E. A. PEEL
C. W. VALENTINE
S. WISEMAN
HELEN M. WODEHOUSE

AND THE FOLLOWING REPRESENTATIVES
OF OTHER COUNTRIES

L. CARMICHAEL
ARNOLD GESELL
I. MEYERSON
JEAN PIAGET
FRED J. SCHONELL
L. M. TERMAN

ASSISTANT EDITOR
L. B. BIRCH

Published by METHUEN AND Co., LTD.

Price Ten Shillings

E. & S. LIVINGSTONE, LTD.

THE YOUNG HANDICAPPED CHILD
By AGATHA H. BOWLEY, Ph.D., F.B.Psy.S.
136 pages. 36 illustrations. 10s. 6d.

THE NATURAL DEVELOPMENT OF THE CHILD
Fourth Edition. By AGATHA H. BOWLEY.
224 pages. 84 illustrations. 10s. 6d.

Also by Dr. A. H. Bowley

Child Care: A Handbook on the Care of the Child Deprived of a Normal Home Life	10s. 6d.
Everyday Problems of the School Child	7s. 6d.
The Psychology of the Unwanted Child	6s.
The Problems of Family Life. *Second Edition.*	6s.
The Spiritual Development of the Child With the Rev. Michael Townroe.	6s.

THE EARLIEST STAGES OF DELINQUENCY
By H. EDELSTON, M.D., D.P.M. 208 pages. 10s. 6d.

JUNIORS
By G. H. PUMPHREY. 180 pages. 15 illustrations. 8s. 6d.

TEVIOT PLACE, EDINBURGH

BOOKS on Psychology—

and allied subjects supplied from stock, or obtained to order. Foreign books not in stock obtained to order. Catalogues on request. Please state interests.

SECOND-HAND DEPARTMENT
Select stock of recent editions. Large and small collections bought.

LENDING LIBRARY
Annual Subscription
from £1 17s. 6d.

Prospectus on request, post free

H. K. LEWIS & Co. Ltd.
136 GOWER STREET
LONDON, W.C.1
Telephone: EUSton 4282 (7 lines)
Business hours: 9 a.m. to 5-30 p.m., Saturdays to 1 p.m

New books from Methuen

THE PSYCHOLOGY OF LEARNING
B. R. BUGELSKI

An appraisal of the current state of learning theory, with particular emphasis on the related field of motivation. The author, who is Professor of Psychology at the University of Buffalo, pays special attention to the problem of past experience and its influence on the future. 50s. net.

THE SCIENTIFIC STUDY OF SOCIAL BEHAVIOUR
MICHAEL ARGYLE

An up-to-date account of psychological research into human social behaviour. There are chapters on interaction between pairs of people, behaviour in small social groups, and human relations in industry.

The author is lecturer in Social Psychology in the University of Oxford.
Methuen's Manuals of Psychology 20s. net.

Methuen & Co. Ltd., 36, Essex Street, London, W.C.2

CONTENTS

PAGE

D. E. Berlyne: Recent Developments in Piaget's Work . 1

J. D. Carsley: The Interests of Children (aged 10-11) in Books 13

M. Gallagher: The Prospect of Marriage: A Study of the Attitudes towards Further Education of a Sample Group of Secondary Technical and Secondary Grammar School Leavers . . 24

M. L. Kellmer Pringle: Differences between Schools for the Maladjusted and Ordinary Boarding Schools . . . 29

D. A. Pidgeon and A. Yates: Symposium: The Use of Essays in Selection at 11+. IV.—Experimental Inquiries into the Use of Essay-Type English Papers . 37

M. A. Davidson, R. G. McInnes and R. W. Parnell: The Distribution of Personality Traits in Seven-Year-Old Children: A Combined Psychological, Psychiatric and Somatotype Study . 48

R. Lynn: Temperamental Characteristics Related to Disparity of Attainment in Reading and Arithmetic 62

Summary of Research Reported in Degree Theses 68

Book Reviews 70

Other Publications Received 74

J. Hemming: Symposium: The Development of Children's Moral Values. 1.—Some Aspects of Moral Development in a Changing Society . . 77

E. Bene: The Objective Use of a Projective Technique, Illustrated by a Study of the Differences in Attitudes between Pupils of Grammar Schools and of Secondary Modern Schools . 89

T. Lee: On the Relation between the School Journey and Social and Emotional Adjustment in Rural Infant Children 101

J. D. C. Robertson: An Analysis of the Views of Supervisors on the Attributes of Successful Graduate Student Teachers. 115

K. M. Miller: Einstellung Rigidity, Intelligence and Teaching Methods 127

M. L. Kellmer Pringle and M. D. Neale: A Note on the use of the Schonell and Gates Reading Tests in the First Year of the Junior School 135

D. N. Lawley and A. E. G. Pilliner: Comments on "Essay Marking Experiments: Shorter and Longer Essays," by D. M. Edwards Penfold 142

D. M. Edwards Penfold: Reply to comments on "Essay Marking Experiments" 145

Book Reviews 147

Other Publications Received 151

THE PSYCHOLOGICAL BASIS OF EDUCATION

By E. A. Peel, M.A., Ph.D.

"This is an important book which may mark a turning-point in the teaching of psychology to teachers." *The Times Educational Supplement*

The object of this book is to provide a fairly concise account of the theory and findings of modern psychology that are of particular use in education. Four major topics are discussed ; the nature of learning, the psychology of individual differences, personal development, and the relation of the individual to society. The learning theories of Hull and Tolman are selected for special consideration.

CONTENTS : Educational Psychology and the Teacher—Early Stages of Learning—Later Theories of Learning—Learning as Insight—The Learner is Active—Some Applications of Learning Theory—The Structure of the Mind—Intelligence—Individual Differences—The Structure of Emotional Life—Emotional Development—Personality—The Person in Society—School Examinations—Record Cards and their Uses.

Price **20s.** net

OLIVER & BOYD: TWEEDDALE COURT
EDINBURGH 1

VOLUME XXVII. FEBRUARY 1957 PART I

RECENT DEVELOPMENTS IN PIAGET'S WORK

BY D. E. BERLYNE
(*Department of Psychology, University of Aberdeen*).*

Jean Piaget's post-war writings are of the utmost importance and interest to educational and child psychologists. But, even when translated, they are very difficult to follow. Dr. Berlyne kindly agreed to supply the following survey of his work on the development of children's psychological functions from birth to adolescence.

—EDITOR.

I.—INTRODUCTION.

PIAGET is known to English-speaking psychologists mainly for his early writings, with their thought-provoking but, according to some critics, disputable accounts of the quaint notions of young children. Doubts have been expressed about the validity of the method of interrogation used for these studies and about the generality of the findings. Repetitions with other populations have not always produced the results that Piaget's works would lead one to expect. At least one writer was moved to dismiss his 'subjective approaches to the analysis of child behaviour' as 'little removed from ordinary literary speculation.'†

Since the 1930s, however, Piaget's researches have been undergoing some gradual but profound changes. He has been turning to more exact and behaviouristic methods of collecting data: close observation of infants, setting older children practical tasks or putting precise questions to them about events enacted in front of them, and psychophysical experiments with both child and adult subjects. His theory has become more detailed and more ambitious in scope, drawing on his knowledge of biology, logic and history of science, all of them fields to which he has contributed. These developments can be summed up by saying that he has changed from one of the most celebrated *developmental* psychologists into one of the most important of contemporary *general* psychologists. But this does not mean that his work has lost any of its importance for those faced with the practical problems of childhood in their everyday work.

Like most contemporary psychologists, Piaget starts from the biological concept of 'adaptation.' He sees adaptation as an interplay of two complementary processes, which he calls '*assimilation*' and '*accommodation*.' Assimilation occurs when an organism uses something in its environment for some activity which is already part of its repertoire. At the physiological level, it is exemplified by the ingestion of food, and, at the psychological level, it embraces a variety of phenomena. Piaget sees assimilation at work, for example, whenever a situation evokes a particular pattern of behaviour because it resembles situations that have evoked it in the past, whenever something new is perceived or conceived in terms of something familiar, whenever anything is invested with value or emotional importance. Accommodation, on the other hand, means the addition of new activities to an organism's repertoire or the modification of old activities in response to the impact of environmental events.

Psychologists accustomed to other conceptual schemes may wonder whether it really helps to group together such multifarious processes under the same rubrics. Is the role played by a cow which a child confuses with a horse really analogous to that played by a cow appearing as roast beef on the child's dinner plate? Although Piaget discusses assimilation and accommodation at great length, some readers may feel that the concepts need to be analyzed more minutely before they can yield unequivocal predictions rather than describing facts already discovered. At all events, assimilation seems to include what learning theorists call ' generalization ' and ' discrimination,' processes determining which response a particular stimulus will elicit, while accommodation covers ' differentiation of responses ' and the learning of new responses.

---

* Now temporarily at the Center for Advanced Study in the Behavioral Sciences. Stanford, California, U.S.A.

† PRATT, K. C.: The Neonate, in Murchison, C. (ed.). *A Handbook of Child Psychology*. Worcester, Mass.: Clark Univ. Press, 1933.

As the child's development proceeds, a more and more complete balance and synthesis between assimilation and accommodation is achieved. The child is able to take account of stimuli more and more remote from him in space and time, and to resort to more and more composite and indirect methods of solving problems.

Piaget agrees with many other theorists in distinguishing ' affective ' and ' cognitive ' factors. The former release energy, while the latter determine how the energy will be applied. Piaget's writings have concentrated on the ' cognitive ' aspect of behaviour rather than on motivation and emotion, but he insists that neither aspect must be overlooked. The child does not undergo separate intellectual and emotional developments. The most dispassionate pursuit of knowledge must be driven by some motive, and the directions in which drives and emotions impel behaviour must depend on the structures made available by the growth of intelligence.

II.—THE PERIOD OF SENSORI-MOTOR INTELLIGENCE (BIRTH TO TWO YEARS.)*

During his first two years, the child gradually advances towards the highest degree of intelligence that is possible without language and other symbolic functions. He begins life with innate reflexes, but these are, from the start, modified and made more effective by learning. New responses are soon acquired, and then complex solutions to problems are achieved by piecing together familiar responses in novel combinations. By the end of the second year, the first signs of the human capacity for symbolization appear : he invents new patterns of behaviour which show him to be representing the results of his actions to himself before they occur. In short, the sensori-motor period sees attainments comparable to the highest found in sub-human animals.

This growing ingenuity in the face of practical problems goes hand in hand with the formation of a less ' egocentric ' and more ' objective ' conception of the world. For some weeks after birth, the world must consist of a succession of visual patterns, sounds and other sensations. The infant comes naturally to pay attention to those external events which are associated with satisfactions or which are brought about by his own actions. Gradually, he builds up a view of the world as a collection of objects continuing to exist even when they are out of his sight and generally preserving the same sizes and shapes, despite the changes in their appearance that come with changes in position. Whereas no distinction between himself and what is outside him can have any meaning for him at first, he comes to conceive of himself as one object among the many that people the world, most of them unaffected by his activities.

The concept of an *object* is bound up with objective notions of *space*, *time* and *causality*, which the child does not possess as part of his native endowment but has to build up gradually through interaction with the world. After learning to select appropriate spatial directions and temporal successions for his actions, he comes to respond to the positions and times of occurrence of events outside himself, using his own body and his own actions as reference-points. Finally, he conceives of a space and a time in which both he himself and external objects are located. He learns, for example, to distinguish occasions when objects are moving independently of him from occasions when they merely appear to be changing positions because he is moving among them. Similarly, he progresses from an understanding of the relationship between his responses and their consequences to an understanding of the causal influence inanimate objects can exert on one another and even on him.

---

* Cf. Bibilography, items 3, 4 and 6 (Ch. IV).

III.—THE ORIGIN OF SYMBOLIC PROCESSES.*

Anything the child has achieved during the sensori-motor period is dwarfed by the prospects introduced by signs and symbols, particularly words and images. They expose him to a world of real and imaginary entities extending far beyond his momentary range of vision or even his life-span. It is a stable and consistent world, whereas the objects he perceives come and go.

Piaget deprecates the long-established belief that images are mere reactivations of traces of past experiences, passively registered by the nervous system. He insists that imagery is an extremely complex and active process, as can be seen from the time it takes to appear after birth. It grows out of the child's imitative capacities and is, in fact, 'internalized imitation.' The gradual extension of imitation during the sensori-motor period proceeds from a tendency to reproduce sounds and visual effects which have just been produced by the infant himself or by somebody else to an ability to copy an increasing range of new responses from an increasing range of models. It reaches its climax and the point at which it can perform symbolic functions with 'deferred imitation,' the imitation of the behaviour of an absent person of whom the child is 'reminded.'

Inanimate objects also can evoke imitation, as, for example, when a child opens his mouth on finding it difficult to open a match-box. Imagery consists of just such symbolic imitation 'internalized' i.e., so reduced in scale that only the subject himself is aware of it. It consists, in other words, of what behaviourists call 'implicit' or 'fractional' responses. When the first indications of imagery emerge about the middle of the second year, the child is beginning, significantly enough, to turn from 'practice' games, in which pleasure is derived from exercising simple activities, to 'symbolic games,' which involve make-believe or role-playing. The child understands, however, the nature of the relation between a symbol and what it signifies ; he knows that the doll is not really a baby or that he is not really a cowboy.

Having learned to use actions and images as symbols and having by now acquired a sufficient vocabulary, he finds himself using words in a similar way. But words, more than images, are responsible for the progressive socialization of thought. Words and the concepts corresponding to them are taken over from the social group. They are, therefore, bound to edge the child's thoughts into line with those of other persons. He can influence and be influenced by, benefit from or suffer from, the beliefs and values of other members of his group and so arrive at an equilibrium and harmony with his social as well as his physical environment.

---

* Bibliography, items 5, 6 (Ch. V) and 16.

IV.—RELATIONS BETWEEN PERCEPTION AND THOUGHT.*

In recent years, Piaget has been spending a great deal of time, together with Lambercier and other collaborators, on the painstaking investigation of visual illusions and related phenomena. This area of research, a time-honoured preserve of the more prosaic type of experimental psychology, may seem remote from the work for which he is best known. It has, nevertheless, given rise to some of his most original and comprehensive ideas, forming the kernel of his whole theory of intellectual functions. Whereas writers influenced by Gestalt psychology or by certain trends in American social psychology have tended to lump all ' cognitive ' processes together, Piaget finds the differences between perceptual and conceptual processes illuminating.

There are two obvious ways in which perception contrasts with thought. One arises from the fact, emphasized by the Gestalt school, that the perceived properties of a stimulus vary according to the pattern of which it is a component. The concepts participating in thought do not share this instability. The essential nature of a number does not change, no matter what the structure into which a mathematician fits it. A journey between two towns may seem longer or shorter in different circumstances, but the distance separating the towns according to our knowledge or our calculations does not fluctuate.

Secondly, perceptions are notoriously variable from person to person and from moment to moment. If we take 1,000 subjects, show them a line three inches long and another two inches long, and ask them to select a third line equal in length to the two combined, we shall expect a distribution of results with a high variance. We shall even expect each subject's response to vary from trial to trial, especially if the two lines are shown in different arrangements. On the other hand, if we take the same 1,000 subjects, show them the figure 2 and the figure 3, and ask them to select a third figure, equal to the sum of the two, the uniformity of the responses will be remarkable.

These differences can be traced back to two related factors which inevitably distort all perception. First, perception is always ' centred ' (*centré*). Sense-organs have to be oriented in one direction at once, and the optical apparatus in particular is so constructed that the centre of the visual field is seen more clearly and in more detail than other parts. As some of Piaget's psychophysical experiments show, the size of a fixated object is over-estimated in comparison with the sizes of peripheral objects. The various parts of the visual field expand and shrink in turn as the gaze wanders from one point to another. The second source of error is the fact that larger portions of a figure are likely to catch the eye more often than others, with the result that the distortions that arise when they are the centre of attention play a disproportionately large part in the net impression of the figure. What we have is, in fact, a biased sample of all possible fixations. From these assumptions, Piaget has derived a formula predicting the direction and extent of ' primary ' visual illusions, i.e., those which are found in infants and lower animals as much as, if not more than, in adult human beings and which can be ascribed to the inherently ' probabilistic ' nature of perception.

Perception is analogous to certain processes in physics, notably in statistical mechanics, which are likewise governed by probability. These processes are irreversible, since they always lead from a less probable to a more probable state. For example, when a hot body is brought into contact with a cool body, heat is transmitted from the former to the latter and not *vice versa*. A spoonful of sugar diffuses evenly through a cupful of tea, but particles of sugar in a mixture do not forgather at one spot. Similarly, the distortions to which perceived figures are subject work predominantly in one direction. They cannot be relied on to balance out.

---

* Bibliography, items 6 (Ch. III) and 14.

Thinking can escape from these limitations, because it is comparable with physical systems of a different type, namely those possessing *reversibility*. An example is a balance with equal weights in the two pans. The depression of one pan is followed by an upward swing which restores the original situation. Such systems are in stable equilibrium precisely because a change can be cancelled by an equal change in the opposite direction. A balance, however, is inflexible in the sense that there is one state to which it invariably reverts. Thought processes require structures which permit of more mobility without threatening disequilibrium. They must be free to flit rapidly from one idea to another and to arrange ideas in new combinations. But systems of concepts must preserve their organization, despite this mobility, if thoughts are to be consistent and if they are to produce a stable conception of the world. The 'dynamic equilibrium' which Piaget attributes to thought can perhaps best be compared with that of a lift and its counterweight. The lift can move freely up and down, and the system remains intact and in equilibrium, no matter what floor is reached. This is because of its reversibility: any movement of the lift is compensated by an equal and opposite movement of the counterweight, and it can also be nullified by an equal and opposite movement of the lift.

The reversibility of logical thought is acclaimed by Piaget as the acme in which the growth of intelligence culminates. The spoken word and the performed action can never be recalled. The influence of something which has been perceived and then disappears from view lingers to infect subsequent perceptions. But a thought can be entertained and then unthought, and everything is as if it had never occurred. We are consequently able to conceive possible solutions for problems which it would be costly, dangerous or impossible to test by action. And no matter how extravagant an idea is considered and then rejected, the coherence of conceptual systems is not threatened. The world represented by thought, unlike that presented by perception, is relatively free from 'centring' (*centration*). It does not change with the location of the thinker or the direction of his attention.

These contrary characteristics are found in a pure form only in the naive perception of the infant on the one hand and in the rigorous thought of the scientist, mathematician or logician on the other. The principal merit of this part of Piaget's work, as far as child psychology is concerned, is the light it sheds on certain processes forming compromises between perception and thought. As we shall see when we return to the chronological sequence, the first attempts at thinking are still contaminated with the short-comings of perception. And perception, after the first months of life, is usually accompanied by 'perceptual activities,' which mitigate its imperfections. There is no way of removing distortion completely from perception, but one distortion can be set against another. The focus of attention can be systematically varied, so that information from a succession of fixations is compared and collated to yield something approaching an objective impression. What appears from one point of view can be related to the perseveration or anticipation of what has been or will be seen from a different angle. 'Perceptual activities' thus contribute to the 'decentring' (*decentration*) of perception and the achievement of 'semi-reversibility,' so called because errors are not corrected exactly but merely tend to cancel out in the long run. Although these activities generally enhance accuracy of perception, they can on occasion lead to 'secondary illusions,' which are less pronounced in younger than in older children. An example is the 'size-weight illusion,' which makes a small object seem heavier than a larger one of equal weight.

V.—The period of pre-conceptual thought (two to four years).*

Before his use of symbolic processes can reach fruition, the child has to re-learn on a conceptual level some of the lessons he has already mastered on the sensori-motor level. For instance, he may have learned to recognize transient stimulus-patterns as shifting appearances assumed by enduring objects. But this does not immediately make him at home with the *concept* of an object. Adults are familiar with the concept of a particular *object* (' *this table,*' ' *Socrates*'), with the concept of a *class* (' *all four-legged tables,*' ' *all men* ') and with the relation of *class-membership* which joins them (' *This is a four-legged table,*' ' *Socrates is a man.*'). These underlie our deductive reasoning, since having, for example, placed Socrates in the class of men, we can infer that Socrates has all the properties characteristic of this class.

The three-year-old child still lacks this equipment and has to use something midway between the concept of an object and that of a class, which Piaget calls the ' pre-concept.' On a walk through the woods, for example, he does not know whether he sees a succession of different snails or whether the same snail keeps on re-appearing. The distinction, in fact, means nothing to him ; to him they are all ' snail.' Similar phenomena are, in some hazy way, identified, so that a shadow under a lamp in a room has something to do with the shadows under the trees in the garden. Contrariwise, a person in new clothes may be thought to require a new name.

Unlike adults, who reason either *de*ductively from the general to the particular or *in*ductively from the particular to the general, the child at the pre-conceptual stage reasons *trans*ductively from the particular to the particular. It is a form of argument by analogy : ' A is like B in one respect, therefore A must be like B in other respects.' Transduction may often lead to valid conclusions, e.g., that if Daddy is getting hot water he must be going to shave, since he shaved after getting hot water yesterday. But it will at other times lead the child into errors of a sort said to be common in psychotics but certainly not unknown in intellectual circles.

VI.—The period of intuitive thought (four to seven years).†

When the child's reasoning has overcome these deficiences, other limitations remain, mainly because thought has not yet freed itself from perception and become ' decentred.' Intuitive thought can best be understood from an experiment Piaget is fond of quoting. The child sees some beads being poured out of one glass into a taller and thinner glass. It is made clear to him that all the beads that were in the first glass are now in the second ; none has been added or removed. He is asked whether there are now more or fewer beads in the second glass than there were in the first. The usual answer at this stage is either that there are more (because the level has risen) or that there are fewer (because the second glass is narrower).

To explain such errors, it may be worth asking why we, as adults, are able to avoid them. The first reason is that we are told by our thought processes that the number of objects in a set, if nothing is added or subtracted, must necessarily remain the same. We usually regard our thought processes as more trustworthy than our perceptions whenever the two conflict. At a conjurer's performance, for example, we do not really believe that the rabbit has been created *ex nihilo* or the lady has been sawn in half. The child at the intuitive stage is, on the other hand, still dominated by his perceptions. His conclusions are still at the mercy of the changes resulting from successive ' centrings.' The second reason is that we take into account several aspects of the situation at once or in turn. We can see that the height of the column of beads has increased and that the width has decreased just enough to compensate for the increase in height. But the child focusses on one aspect and overlooks others. In his reasoning as in his perception, ' centring ' causes one element to be overemphasized and others to be relatively ignored. The instructiveness of such examples for adults, who might smile at the child's mistakes in the bead experiment but be liable to precisely the same sort of misjudgment in relation to, say, political or social problems, needs hardly be laboured.

---

* Bibliography, items 5 and 6 (Ch. V).

† Bibliography, items 5 and 6 (Ch. V).

VII.—THE PERIOD OF CONCRETE OPERATIONS (SEVEN TO ELEVEN YEARS).*

We come at last to the first reasoning processes that would satisfy logicians. Logical (or, as Piaget calls it, ' operational ') thought emerges when a certain basic stock of concepts has been acquired and when these concepts have been organized into coherent systems. The concepts which figure in operational thought are called ' operations ' because they are *internalized responses*. They grow out of certain overt actions in exactly the same way as images grow out of imitation. Three sorts in particular are of importance :

(1) *Classes.*—The concept of a ' class ' or the operation of ' classification ' is an internalized version of the action of grouping together objects recognized as similar. Having learned to pick out all the yellow counters in a heap and *place* them together in one spot, the child acquired the ability to *think of* all yellow objects together and thus form the concept of the ' class of all yellow objects.' This means that some part of what happens in the nervous system and musculature when yellow objects are manually gathered together occurs whenever yellow objects are grouped together in thought. Once formed, classes can be joined to form more inclusive classes, so that elaborate systems of classification are built up, the one used by biologists being the clearest illustration.

(2) *Relations.*—Asymmetrical relations, such as 'a is longer than b' or ' x is the father of y,' derive by internalization from *ordering* activities, e.g., from placing objects in a row in order of increasing size. The best example of the complex systems which ordering relations can form is the family tree.

(3) *Numbers.*—The number system is the joint product of classification and ordering. The number 17, for instance, depends on the operation of grouping 17 objects together to form a class and that of placing 17 between 16 and 18 in the sequence of natural numbers.

Systems of operations are called ' groupings ' (*groupements*), and their stability depends on their having five properties. Unless these properties are present, the relations between the elements of a grouping will change as attention is directed to different parts of them, as happens with perceptual patterns, and thought will not be immune from inconsistency. The five properties are as follows :

(1) *Closure.*—Any two operations can be combined to form a third operation (e.g., 2+3=5 ; *all men and all women=all human adults* ; *A is* 2 *miles north of B and B is* 1 *mile north of C=A is* 3 *miles north of C*.

(2) *Reversibility.*—For any operation there is an opposite operation which cancels it (e.g., 2+3=5 but 5-3=2 ; *all men and all women=all human adults*, but *all human adults except women=all men* ; *A is* 2 *miles north of B and B is* 1 *mile north of C=A is* 3 *miles north of C*, but *A is* 3 *miles north of C and C is* 1 *mile south of B=A is* 2 *miles north of B*).

---

* Bibliography, items 6 (Ch. II and Ch. V), 7, 11 and 16.

(3) *Associativity.*—When three operations are to be combined, it does not matter which two are combined first. This is equivalent to the possibility of arriving at the same point by different routes (e.g., (2+3) +4=2+(3+4) ; *all vertebrates and all invertebrates=all human beings and all sub-human animals* ; *a is the uncle of b and b is the father of c=a is the brother of d and d is the grandfather of c*).

(4) *Identity.*—There is a ' null operation ' formed when any operation is combined with its opposite (e.g., 2 − 2=0 ; *all men except those who are men= nobody* ; *I travel* 100 *miles to the north and I travel* 100 *miles to the south=I find myself back where I started*).

(5) The fifth property has two versions, one for classes and relations and the other for numbers :

(*a*) *Tautology.*—A classification or relation which is repeated is not changed. This represents the fact, recognized by logicians but not always by conversationalists, that saying something over and over again does not convey any more information than saying it once (e.g., *all men and all men = all men* ; *a is longer than b and a is longer than b=a is longer than b*).

(*b*) *Iteration.*—A number combined with itself produces a new number (e.g., 3+3=6 ; 3×3=9).*

VIII.—THE PERIOD OF FORMAL OPERATIONS (ELEVEN TO FIFTEEN YEARS).†

The eleven-year-old can apply ' operational ' thinking to practical problems and concrete situations. The adolescent takes the final steps towards complete ' decentring ' and ' reversibility ' by acquiring a capacity for abstract thought. He can be guided by the *form* of an argument or a situation and ignore its *content.* He need no longer confine his attention to what is real. He can consider hypotheses which may or may not be true and work out what would follow if they were true. Not only are the hypothetico-deductive procedures of science, mathematics and logic open to him in consequence but also the role of would-be social reformer. The adolescent's taste for theorizing and criticizing arises from his ability to see the way the world is run as only one out of many possible ways it could be run and to conceive of alternative ways that might be better.

Quite a variety of new intellectual techniques become available at the same time. The most important new equipment of all is the *calculus of propositions.* At the concrete-operations stage, he was able to use the branches of logic known as the *algebra of classes* and the *algebra of relations.* Now he can supplement these with forms of reasoning bearing on the relations between propositions or sentences. Propositional calculus uses ' second-order operations ' or operations on operations. An example would be ' *either sentence p is true or sentence q is true.*' Another would be ' *if sentence r is true, then sentence s must be true* ' or, in the parlance favoured by logicians, ' *r implies s.*'

A large part of Piaget's information on this period comes from Inhelder's ingenious experiments, in which children were invited to discover elementary laws of physics for themselves with the help of simple apparatus. Children at the intuitive-thought stage vary conditions haphazardly and observe what happens in particular cases without deriving any general principles. At the concrete-operations stage, one factor at a time is varied, and its effects are duly noted.

---

* Readers with mathematical interests will notice that, in so far as these properties refer to numbers, they are equivalent to the defining characteristics of a *group.* Groupings of classes and relations, on the other hand, are almost, but not quite, *groups* and almost, but not quite, *lattices.*

† Bibliography, items 7, 9, 10, 11, 15, 17.

Not before the formal-operations stage does the child plan truly scientific investigations, varying the factors in all possible combinations and in a systematic order. The pedagogical implications of Inhelder's work are unmistakeable. Children with no previous instruction appear to be capable of learning scientific laws in this way, with, presumably, more motivation and more understanding than are produced by traditional teaching methods. But, according to Piaget and Inhelder, they are not capable of the sort of thinking that makes use of such laws before the advances of the formal-operations stage have been completed.

Piaget asks why so many new ways of thinking become available about the same time, despite their superficial dissimilarity. It is, he concludes, because they all require systems of operations with similar structures, and the child is not able to organize his thinking in accordance with such structures before adolescence. He has recently been much impressed with the possibilities of modern symbolic logic and certain non-numerical branches of mathematics as means of describing the structures common to apparently different intellectual processes. This is not one of the ways in which logic has usually been used by psychologists in the past; Piaget is interested in using 'logical models' for much the same purpose as other psychologists have begun to use 'mathematical models.'

One new acquisition is the ability to use systems of operations in which each operation has two distinct opposites. A class (e.g., '*all vertebrate animals*') has the sort of opposite called an *inverse* ('*all invertebrate animals*'). A relation (e.g., '*a is twice as heavy as b*') has a *reciprocal* ('*b is twice as heavy as a*'). But '*p implies q*' has both an inverse ('*q does not imply p*') and a reciprocal ('*q implies p*'). Likewise, when the adolescent experiments with a balance, he discovers that the effects of one operation (e.g., increasing the weight in the right-hand pan) can be cancelled either by the inverse operation (reducing the weight in the right-hand pan to its original value) or by the reciprocal operation (increasing the weight in the left-hand pan by the same amount.) Such systems with two opposites have a structure known to mathematicians as the '*four group*.'

The four group can be shown to provide the operations necessary for dealing with *proportionality*. It is no accident that the laws governing equilibrium between weights in the pans of a balance are understood at about the same age as the laws governing the sizes of shadows. In one of Inhelder's experiments, the subject is given two vertical rings of different diameters and has to place the rings between a candle and a screen in such a way that their shadows will coincide. Adolescents discover that the problem is solved when the ratio between the distances of the two rings from the candle is the same as the ratio between their diameters. Understanding proportionality opens the way to understanding *probability*, since, when we speak of the probability of a six in a game of dice, we mean the proportion of throws that will produce sixes in the long run.

Combinatorial analysis, depending on the structures mathematicians call '*lattices*,' is another equally fruitful new attainment. Suppose that we have two ways of dividing up animals—into 'vertebrates (*V*)' and 'invertebrates (*v*)' and into 'flying (*F*)' and 'non-flying (*f*).' A child at the concrete-operations stage is capable of allotting a particular animal to one of the four possible classes, (*V.F.*), (*V.f.*), (*v.F.*) and (*v.f.*). An adolescent at the formal-operations stage is capable of going further and considering all the sorts of animals that there are in the world or the sorts there conceivably could be. There are now *sixteen* possibilities: there might be no animals at all, there might be animals of all four classes, there might be (*v.F.*) only, there might be (*V.F.*), (*V.f.*) and (*v.f.*) animals but no (*v.F.*), etc. Now each of these sixteen combinations corresponds to one of the sixteen relations between two propositions recognized by modern logic. For example, '*if an animal can fly, it must be a vertebrate*' would correspond to (*V.F.*) *or* (*V.f.*) *or* (*v.f.*), i.e., the (*v.F.*) possibility is excluded. We can understand, therefore, why permutations and combinations and complex logical relations are mastered more or less simultaneously.

The mastery of logical relations between propositions is well illustrated in Inhelder's experiments. All attempts to study the relations between the phenomena of nature, whether in the laboratory or in practical life, must use them : '*If I put the kettle on the stove and light the gas, the water will boil*' ; '*It will rain or snow tomorrow unless the forecast was wrong or unless I read a description of today's weather and thought it was the forecast for tomorrow,*' etc. The ability to think in terms of all possible combinations, which appears together with the ability to use complex statements like these, is clearly revealed when adolescents are set one of Inhelder's most instructive problems. Five vessels, all containing colourless liquids, are provided ; A, B and C, when mixed, will turn pink, D will remove the colour, and E will have no effect. The properties of the liquids can be discovered only by systematically examining mixtures of every possible pair, every possible trio, etc., in turn.

IX.—AFFECTIVE DEVELOPMENT.*

The child's physiological constitution makes him liable, right from birth, to emotional and drive states. These pleasant and unpleasant states come to be aroused, through some sort of conditioning, by the external stimulus patterns which regularly accompany them, and, when he had learned to perceive in terms of objects, he comes to like or dislike these. Human beings are naturally more important sources of satisfaction and distress than other objects, and so their actions and they themselves will have especially strong positive and negative values attached to them.

The social influences to which the appearance of language and other symbols makes the child amenable are manifested particularly clearly in the formation of 'inter-individual feelings.' The ability to picture how the world looks from another person's point of view includes the power to represent to oneself the feelings aroused in him by one's own actions. The child takes over other people's evaluations of his own behaviour and builds up an attitude to himself derived from his estimates of their attitudes to him. The stage is then set, during the pre-conceptual and intuitive periods, for the first moral feelings. These take the form of a belief in absolute prohibitions and prescriptions, derived from parental orders but somehow enjoying an existence and validity in their own right. Acts are felt to deserve punishment according to how far they depart from what is permitted, without reference to intentions or other mitigating circumstances.

When he reaches the period of concrete operations, the child can form groupings of values, as of other classifications and orderings. He can systematize his values according to their relative priorities and their mutual affinities, so that his evaluations and his motives may be consistent with one another. He can subordinate his actions to future needs, thereby achieving that 'decentring' from the present which we call *will.* His addiction to 'games with rules,' which replace 'symbolic games' about this time, shows him to have arrived at a less primitive conception of moral rules. He now sees them as conventions, accepted by a social group for the benefit of all, capable of being changed by common consent, and arising out of mutual respect between equals.

---

* Bibliography, items 6 (Ch. VI) and 13.

By the end of the formal-operations stage, feelings become 'decentred' still further, as they are released from the domination of what is known to be actually true. Motivation and evaluation now depend on *ideals*, and everything tends to be judged by how far it approximates to or falls short of the theoretical states of affairs that would fulfil these ideals. The adolescent views his own activities and plans as part of the total activity of the social group. He begins to think of himself as a fully fledged member of society, free to imitate or criticise adults. With the 'decentring' which implants the individual in the community and subordinates his activities to collective goals, the formation of the personality is complete.

X.—CONCLUSIONS.

It is evident that Piaget's latest work will not silence his critics altogether. He still does not pay much attention to questions of sampling. Some projects, e.g., Inhelder's on adolescents, seem to have used a large part of the school population of Geneva. The data on the sensori-motor period, on the other hand, come mainly from observation of Piaget's own three children, hardly the children of the Average Man! But Piaget might well retort, like Kinsey, that such bodies of data, however imperfect, are all we have of comparable density.

Except for some means and mean deviations in his reports of perceptual experiments, he provides few statistics. There are generally no measures of variance, which one suspects must be considerable, no tests of significance, just a categorical statement that at such and such an age children do such and such, with a few specific illustrations. He is not much affected by the growing vogue for rigorous theories, with precise statement of assumptions, derivation of predictions and operational definition of concepts.

Be that as it may, Piaget is, without any doubt, one of the outstanding figures in contemporary psychology, and his contributions will eventually have to be reckoned with much more than they are both in the management of children and in many areas which may not seem directly connected with child psychology. His ideas are closely tied to observation of behaviour, and this makes them the sort of psychology which moves science forward because it is testable by reference to the facts of behaviour. At the same time, it goes beyond the facts just sufficiently to open up new lines of research and to attempt the sort of synthesis which is one of the chief aims of science.

Not the least reason for paying attention to Piaget's work is the relation it bears to trends followed by English-speaking psychologists. At times, his conclusions parallel those reached independently by other investigators; at other times, they serve to correct or supplement what psychologists with other approaches have to say. Like those influenced by Gestalt psychology, Piaget affirms that perceptions and thoughts cannot be understood without reference to the wholes in which they are organized. He disagrees with them in denying that wholes are unanalyzable into component relations and in insisting that the wholes figuring in thought are radically different from those figuring in perception. There are, throughout his writings, many reminders of psycho-analytic concepts—the 'omnipotence' and 'oceanic feeling' of infancy, 'functional pleasure,' the formation of the ego and the super-ego, the advance from the pleasure principle to the reality principle. But he makes many detailed criticisms of psycho-analytic theories, and the child as described by him certainly seems tranquil and studious by comparison with the passion-torn 'polymorphous pervert' that emerges from Freudian writings.

But Piaget's closest affinities are undoubtedly with the neo-behaviourists. He does not hold with early attempts to explain everything by 'conditioned reflexes ' or ' association.' But many of his observations and many aspects of his theory harmonize extremely well with conceptions of learning based on studies of what has come to be called ' instrumental ' or ' operant conditioning.' The sequence of more and more complex behaviour patterns which he depicts as outgrowths of simple reflexes and habits parallels Hull's list of progressively more intricate ' adaptive behaviour mechanisms,' found in animals.* And Piaget's view of images and thought operations as ' internalized ' overt responses approximates very closely to the view prevalent among stimulus-response learning theorists.

One body of work which has grown up in Great Britain and the U.S.A. and which Piaget is eagerly endeavouring to bring into relation with his own findings is that centering on cybernetics, information theory and game theory.† But it is to be hoped that other common ground between his psychology and others with different starting-points will be explored. It is certainly high time that the national self-sufficiences which disfigure psychology in contradistinction to other branches of science were left behind.

XI.—SELECTED BIBLIOGRAPHY.

A.—IN ENGLISH.

1 MAYS, W. (1955). How we form conceps. *Sci. News*, 35, 11-23.
2 MAYS, W. (1954). Professor Piaget's épistémologie génétique. *Proc. II. Int. Cong. Phil. Sci.*, 5, 94-99.
3 PIAGET, J. (1953). *The Origin of Intelligence in the Child.* London: Routledge and Kegan Paul.
4 PIAGET, J. (1955). *The Child's Construction of Reality.* London: Routledge and Kegan Paul.
5 PIAGET, J. (1951). *Play, Dreams and Imitation in Childhood.* London: Heinemann.
6 PIAGET, J. (1950). *The Psychology of Intelligence.* London: Routledge and Kegan Paul.
7 PIAGET, J. (1953). *Logic and Psychology.* Manchester: Univ. Press.
8 PIAGET, J. (1952). Genetic psychology and epistemology. *Diogenes*, 1, 49-63.

B.—IN FRENCH.

9 INHELDER, B. (1954). Les attitudes expérimentales de l'enfant et de l'adolescent. *Bull. de Psychol.*, 7, 272-282.
10 INHELDER, B., and PIAGET, J. (1955). *De la logique de l'enfant à la logique de l'adolescent.* Paris: Presses Universitaires de France.
11 PIAGET, J. (1949). *Traité de logique.* Paris: Colin.
12 PIAGET, J. (1950). *Introduction à l'épistémologie génétique.* Tome I: *La pensée mathématique*, Tome II: *La pensée physique*, Tome III: *La pensée biologique, La pensée psychologique, La pensée sociologique.* Paris: Presses Universitaires de France.
13 PIAGET, J. (1953-4). Les relations entre l'intelligence et l'affectivité dans le développement de l'enfant. *Bull. de Psychol.*, 7, passim.
14 PIAGET, J. (1954-5). Le développement de la perception de l'enfant à l'adulte. *Bull. de Psychol.*, 8, *passim.*
15 PIAGET, J. (1954). La période des opérations formelles et le passage de la logique de l'enfant a celle de l'adolescent. *Bull. de Psychol.*, 7, 247-53.
16 PIAGET, J. (1949). Le probléme neurologique de l'intériorisation des actions en opérations réversibles. *Arch. de Psychol.*, 32, 241-58.
17 PIAGET, J. (1953). Structures opérationelles et cybernétique. *Année Psychol.*, 53, 379-88.
18 PIAGET, J. (1954). Les lignes générales de l'épistémologie génétique. *Proc. II. Int. Cong. Phil. Sci.*, 1, 26-45.

---

* HULL, C. L.: *A Behavior System.* (New Haven: Yale Univ. Press, 1952, pp. 347-50.)

† Bibliography, item 17.

(*Manuscript received, May 14th, 1956*).

THE INTERESTS OF CHILDREN (AGED 10-11) IN BOOKS

By J. D. CARSLEY
(University of Liverpool, Institute of Education).

Summary. 1.—An investigation was undertaken in March, 1955, into the interests in books (excluding textbooks) of children aged 10-11.

2.—Children were asked to rate the books they had read at home or in school during the last month. From their replies it was found that in school it was usually a little more enjoyable to listen to the teacher reading aloud than to read silently, and more enjoyable to read silently than to listen to other pupils reading, though this did not apply equally to all social areas. It was noted that a larger amount of time was spent on silent reading in schools in poor neighbourhoods than elsewhere. A classification of books read aloud by the teacher showed that teachers generally followed the established patterns of the accepted children's classics and of popular current children's stories.

3.—It was found that for 95 per cent. of the children, the home provided the best environment for the private reading of books, there being less distraction and interruption than in school. Perhaps as a corollary, it was found that children more frequently gave high ratings to books read at home than to those read in school. On the other hand, a few children stated that they had read no book at home during the last month, and for boys in poor social areas there was, on average, less reading of books at home than at school.

More than half of the children were active members of a public library, except in schools serving housing estates where the proportion was about 1 in 4. Some children had ceased to be active members, being dissatisfied with the library service for various reasons.

4.—*Reading preferences based on stated likings for classes of books showed that* there were more differences between boys and girls than between groups of children from dissimilar social areas, particularly in their attitudes to school stories, to some classes of non-fiction and to poetry.

There was general agreement on the technique followed by children in choosing a book. It showed that children generally preferred to make their own selection, but required time to judge of the contents of a book before doing so. There was a widespread feeling that illustrations tended to spoil the story.

In view of the complex pattern of children's individual likes and dislikes, it would appear necessary to give children of this age the widest possible choice from a variety of children's classics, adult classics, contemporary children's fiction, reference books, non-fiction including biography, and children's annuals.

5.—Films, television and radio have a guiding influence over children's reading by stimulating many of them to read the book after they have seen or heard the dramatised version, and most of the reading reported in this survey was of children's classics. Films gave the greatest stimulus to boys' reading and television to girls.' On the data here available, it is not possible to say whether children spent less time on reading because of the advent of television into the home, but there is evidence to show that where there was a television set in a house the tendency was for children to spend more time in viewing than in reading.

I.—Introduction.

The investigation was undertaken in March, 1955, with the purpose of obtaining information about the interest in books (other than text-books) of junior school children, of the sources from which they obtained books, and about their preferences as shown in the ratings they gave to books considered singly, or their likes and dislikes according to an easily understood subject classification. There has been no published investigation into children's reading interests carried out in English junior schools, despite the importance traditionally placed on the teaching of reading at this stage. Jenkinson's survey* was an enquiry into the reading interests of children in secondary and senior schools between the ages of 12-15, and was not confined, as the present enquiry is, to the field of books. Various surveys have been made of children's reading interests by public librarians, some of which are enumerated by Dunlop.* It was felt that the time for this investigation was opportune in view of the increasing importance placed on the supply of suitable library books in schools, both as a source of experience and information and as a way of raising standards of literacy.

---

* Jenkinson, A. J. (1940). *What do Boys and Girls Read?* Methuen.

A questionnaire was drawn up as a result of several pilot experiments and subsequent discussions with teachers. It was then administered in junior schools by the teachers, each question being read out separately with explanatory comments as suggested in covering notes. Information concerning the supply of library books in the school and the numbers of children in active membership of the public library was supplied by the teachers. Discussions which were later held with groups of six children chosen at random from thirteen classes in six schools provided additional information. These discussions had many common topics, viz., the use of reference books in connection with hobbies and pastimes, the reasons for preferring home reading to school reading, attitudes to the illustration of stories, the attitude of parents to children's reading, the reasons for reading a book several times and for reading a book after the film or television play had been seen, and the adequacy of book-supply. The discussion generally continued from 30 to 45 minutes.

The questionnaire was answered in fifty-five classes in thirty schools from five local education authorities in the Merseyside area (two counties and three county boroughs). A total of 2,040 returns (1,054 boys, 986 girls) was obtained from children aged 10+ on August 1st, 1954. A return was not expected from a child who was not literate enough to answer a question after the teacher had read it aloud. No figures are available to show what proportion of children aged 10+ did not complete a questionnaire, although subsequent enquiries in several of the larger schools showed that most children (estimated as between 90-97 per cent. of those present) were able to do so. Schools were selected to give samples of different organisations (mixed and single sex, single-stream and four-stream, county and voluntary) and especially of a variety of social backgrounds so that the findings would not be biased by the group being unrepresentative. A rough classification was made according to the neighbourhood of the school: (*a*) residential (many owner occupiers), (*b*) good working-class—styled in the analysis of returns as Artisan A, (*c*) poor housing conditions and cultural environment—styled Artisan B, (*d*) housing estates—mainly post-war. The line of demarcation is not a firm one, and several of the schools drew on a socially heterogeneous population. Thus, comparisons of the returns from one group with those from another must be interpreted with caution, the safest comparisons being made between the two extremes—the residential and the Artisan B. The returns were analysed according to sex and to the four social categories, and any differences between the sexes and social groupings have been noted when judged to be significant.

II.—BOOKS READ IN SCHOOL.

The questionnaire asked children to rate (3, 2, 1, 0 marks) books they had read in the last month or so, and to state whether the book had been read silently, or aloud by the teacher or pupils.

(*a*) *Books read silently.*—From the teachers' reports it can be stated that the books read silently in school were usually borrowed from the class library

---

* DUNLOP, D. C. (1950). *Children's Leisure-reading Interests.* Scottish Council for Research in Education Studies in Reading, Vol. II, University of London Press.

(54/55 classes had libraries which might contain 50-450 books), or from the school library (11/30 schools had a central collection of books, often purely reference, varying in size from 100-826 books, average 280). The schools obtained their books by purchase from a requisition allowance or a small library grant, or by loan from the public library. Some also received gifts from pupils or P.T.As. and bought books from the proceeds of school sales and social activities.

Books started in school were seldom taken home to read, and children did not often bring their own books to read in school. Thus, silent reading was largely a characteristic school activitiy, with a recognised place on the time table, which might absorb from half-an-hour to four hours of a child's school time weekly. More time on an average was given to silent reading in schools in Housing Estate and Artisan B areas (2¼ hours) than in schools in Residential and Artisan A areas (1¼ hours). In addition, an unspecified amount of time was available to individual children as other work was finished. The class supply of books at times offered only limited choices, as is shown by the low rating given by some children to the books they read: in one class everyone was reading silently the same story.

TABLE 1

		Children's Rating of books read silently in school				Books read per head in the month
		3 marks	2 marks	1 mark	0	
Boys	N 1054	56·1%	31·8%	9·4%	2·6%	2
Girls	N 986	59·3%	29·4% -	9·2%	2·1%	2

A greater interest was shown in school silent reading by children living in poor areas than by children living in Residential areas (69 per cent. of the books read received the highest rating in Artisan B areas, as against 54 per cent. in the Residential areas). This may be explained partly by the meagre nature of their home libraries and partly by the larger amount of time allotted in school to silent reading. It is also probable that children in Residential areas, more accustomed to reading at leisure for long spells, are less favourably inclined to silent reading in school with its various limitations.

(*b*) *Books read aloud by the teacher.*—In 24/30 schools there was reading aloud by the teacher, in 7/30 there was reading aloud by the pupils to the class and in 3/30 there was no reading aloud recorded during the month.

TABLE 2

		Children's Ratings of books read aloud by the teacher			
		3 marks	2 marks	1 mark	0
Boys	N 788	64·8%	29·3%	5·3%	·7%
Girls	N 751	61·1%	28·6%	8·4%	·2%

It is evident that it was usually rather more enjoyable to listen to the teacher reading than to read silently or to listen to other pupils reading (see Table III). The practice of reading aloud by the teacher has as its main purpose the enjoyment together of a good story or poem, whether rather difficult classics are presented at length or simpler stories are introduced to the children. In most cases, the purpose seemed to have been achieved. It should be noted that boys in Artisan A areas found it especially enjoyable (74 per cent. of books received a maximum rating as against 54 per cent. for silent reading).

Books read aloud by the teacher may be classified as follows:

(1) Classics (using the term loosely of books with enough appeal to survive).

(i) Children's.

Alice in Wonderland.
Black Beauty.
Emil and the Detectives.
Folk Tales (Steedman).
Heidi.
The Heroes.
Huckleberry Finn.
The Jungle Book.
Just so Stories.
Kidnapped.
King Arthur.
Lamb's Tales.
Little Women.
Pinnochio.
Swiss Family Robinson.
Tanglewood Tales.
Tom Sawyer.
Treasure Island.
Wind in the Willows.
The Water Babies.

(ii) Adult books (often in abridged or adapted versions).

Buchan—Greenmantle, Thirty-nine Steps.
Dickens—Christmas Carol, David Copperfield, Oliver Twist, A Tale of Two Cities.
Other books:
Count of Monte Cristo.
Don Quixote.
Lorna Doone.
Moby Dick.
Three Men in a Boat.
Uncle Tom's Cabin.

(2) More recent books.

(i) Children's.

Various books by: Enid Blyton, R. Crompton, W. E. Johns, Arthur Ransome.
Gibson: Wreckers of Pengarth.
Masefield: Jim Davis.
Batten: Redruff.
Whitehouse: Wings of Adventure.
Seth-Smith: Zoo Man.

(ii) Adults.
Hunt: Ascent of Everest.
Williams: The Wooden Horse.

The list shows that teachers generally aimed to strike a balance between the claims of accepted books and those of living writers. Their choice of modern writers for children showed that on the whole they were following the generally established pattern of popular fiction rather than seeking to widen horizons by introducing stories hardly known to the children but highly praised by the critics in periodicals like the *School Librarian, Junior Bookshelf,* and *The Times Literary Supplement,* and included in selective annotated book lists published by the Library Association and the National Book League.

TABLE 3

		Children's ratings of books read aloud by children.			
		3 marks	2 marks	1 mark	0
Boys	N 121	48·5%	31·8%	17%	2·7%
Girls	N 324	51·6%	33·3%	10%	5·2%

(*c*) *Books read aloud by children.*—Less than 10 per cent. of the total entries of books read in school were of books read aloud by children, and of these 8 per cent. were made by girls. Only in a few classes (12/55) was this form of reading a regular feature and then generally because the head teacher or teacher was desirous that children should acquire the social skill of being able to read aloud clearly and expressively. When this means, as it sometimes does, that less proficient readers take their turn in reading to the whole class the enjoyment of the story is diminished both for reader and listener. This may explain the relatively low proportion of books given maximum ratings by children in Residential areas (36·6 per cent.).

III.—READING AT HOME.

The questionnaire asked the child: (*a*) to list and rate books read in about the last month, (*b*) to state the source of supply, (*c*) to state the number of books in his possession at home, (*d*) to say where he most enjoyed reading a book.

TABLE 4

		3 marks	2 marks	1 mark	0
Boys	N 1033	75·4%	20·1%	3·6%	1%
Girls	N 975	73·1%	20·7%	4/6%·	1·5%

Area	Number of books read at home in about one month	
	Boys	Girls
Residential ..	3·8	4·2
Artisan A ..	2·5	3·8
Artisan B ..	1·4	2·6
Housing Estate	2·6	3·4

(*a*) *Children's Ratings of Books.*—Books read at home were given a higher percentage of maximum ratings than those read in school by all groups except Artisan B (girls). In answer to the questionnaire, 95 per cent. of the children stated that the home provided the best environment for silent reading, 90 per cent. of the children in Artisan B areas being of this opinion. About 60 per cent. of the children stated a preference for reading in bed. In discussion, children gave reasons for preferring home to school reading: at home there was less distraction of attention by noise and movement, less likelihood of interruption, longer spells could be spent with a book, and there was greater comfort and a more complete feeling of relaxation. Dissatisfaction with the choice of books available in school was not often given as a reason for preferring home reading.

Most children read again books they had liked on first reading, some as often as five times, as was shown in discussion and from lists of home reading which several classes of children had prepared. The reasons given for re-reading were: to revive memories of the story, to meet again its familiar characters, and occasionally because there was no fresh reading at hand.

The small amount of home reading credited to the average literate or semi-literate boy in the Artisan B areas suggested that here is a problem which requires (and often receives) special treatment. Discussion with children in this type of area revealed that the parents' attitude was not generally a discouraging one. Children who had found some difficulty in learning to read reported that their mothers had read with them, either from a copy of the school reader which had been bought or from " one of mum's books." The illiterate or barely semi literate child does not come within the scope of this enquiry, and a higher proportion of them may be expected in these areas than elsewhere. Table VI shows that home libraries are often pitiably meagre with even the more literate children in these areas. Where poverty of reading experience is shown to exist mainly through lack of books at home, it is argued by some teachers that schools in these areas should be given additional library grants to help them redress the balance and so to increase standards of literacy. The books so purchased should be made available for home reading, since it is here that the striking contrast is discernible in reading habits.

TABLE 5

	Public Library	Gifts and prizes	Bought	Borrowed from a friend	Home	School or class library
Boys N 1033	19·1%	37·8%	8·6%	10·2%	21·7%	2·7%
Girls N 975	19%	36·7%	6%	11·7%	23·2%	3·3%

NOTE.—These figures represent percentages of books from each source of supply.

(*b*) *Sources of books read at home.*—The public library was most used by boys in Residential areas and by girls in Artisan B areas (both 24 per cent) and least used by boys and girls on Housing Estates (10 per cent.). The school or class library was most used for home borrowing by girls in Artisan B areas (5 per cent.). Returns by the teachers showed that some children did not continue their membership because of a long or difficult journey to the nearest library (sometimes given as 1½ miles). Children said they lost interest because of having to pay fines or having to account for the damage done to a book by a younger member of the family. Other occasional children's criticisms were that the best books were generally out, that silence was insisted upon, and that there was not a place to sit and dip into a book before deciding to borrow it

Returns by teachers showed that active membership of the public libraries was: Residential areas 62 per cent., Artisan A areas 49·7 per cent., Housing Estate 26·6 per cent., Artisan B 52 per cent.

Some schools do permit home borrowing, but most teachers are unwilling to risk depleting their slender stock which is required for silent reading periods in school. There is an increasing trend for schools to have public library books on loan, and these are the books which are often available for borrowing.

TABLE 6

MEDIAN NUMBER OF BOOKS (INCLUDING ANNUALS) BELONGING TO CHILDREN.

Area	Boys	Girls	
Residential ..	45	47	Range in size of home libraries 0–400
Artisan A ..	21	27	
Artisan B ..	9	12	
Housing estate	21	21	

Discussion with children showed that some children in Residential areas were given book tokens by relations which involved them in a visit to a city bookseller's. They then made their selection from the books in the shop, generally without any previous idea of what book or class of book they wanted. Most book-gifts they received were purchased at local newsagents' and were not representative of the full field of children's books.

(*c*) *Children's Home Libraries.*— Children were asked to count their own books several days before the questionnaire was given. The numbers of books belonging to children in C and D streams were notably lower than those belonging to children in A and B streams. Lists of home libraries prepared by the children in several schools showed that annuals were predominant in many collections and that children tended to receive in succeeding years the current copy of the annual they first received at seven years of age. The annual was shown to be a slightly more popular present than a long story among boys in Artisan A and B areas or in ' C ' and ' D ' streams.

IV.—CHILDREN'S BOOK PREFERENCES.

The questionnaire asked the child : (*a*) to indicate, by underlining, classes of books for which he had a marked liking or the opposite, (*b*) to indicate by a tick or cross what was the strongest and what the weakest factor influencing him in choosing a book, (*c*) to name his favourite book, and (*d*) to name a book he had disliked reading (if any).

TABLE 7

Boys' (N 1030) Votes	Position	Girls' (N 972) Votes	Position	Category	Total Votes	Position
722	1	764	1	Adventure stories ..	1486	1
668	2	654	2	Mystery stories	1322	2
405	4	228	5	*Encyclopedias	633	3
438	3	163	8	*Sport	601	4
183	6	343	4	*Biblical stories	526	5
—47	13	361	3	School stories	314	6
185	5	115	10	Humorous stories ..	300	7
178	7	115	10	Legends	293	8
107	10	123	9	*How to do things ..	230	9
167	9	51	13	Historical stories ..	218	10
78	11	107	12	Animal stories ..	185	11
—79	15	165	7	*Nature study	86	12
171	8	—248	15	*How things work	—77	13
—312	17	185	6	*Poetry	—127	14
11	12	—141	14	*Travel	—130	15
—52	14	—487	18	Space travel stories	—539	16
—197	16	—357	17	Cowboy stories ..	—554	17
—744	18	—350	16	Fairy stories	—1094	18

* Non-fiction.

(*a*) *Children's Book Preferences according to stated categories.*—The final vote was arrived at by counting a preference as a plus and a dislike as a minus.

There were some differences between the areas. The most notable deviation was in the Artisan B area boys, where positive votes were given for Space Travel (+2), for Poetry (+10), and negative for Animals (—9), How Things Work (—6), How to do Things (—8), Travel (—11). In the case of Residential area boys a positive vote was given for School stories (+15). For girls the differences were : Residential area, a negative vote for Historical stories (—18), and Artisan A area, a negative vote for Nature Study (—2).

But the most notable differences were between the sexes, as the table shows, especially in the case of School stories, some classes of non-fiction and Poetry.

In discussion, children used the word ‘ crazy ’ to describe Cowboy stories, ‘ too far-fetched ” to describe ‘ Space ” fiction, and ‘ babyish ’ for Fairy stories. A high proportion of girls said they enjoyed reading boys' School stories, and some boys said that their dislike of School stories did not spring from a dislike of school. Historical stories appealed to boys because of the battles, fights and pursuits found therein. Biblical stories were preferred when re-told as the language of the Bible caused difficulty. Poetry was only occasionally popular with boys. The commonest reasons for dislike were : compulsory memorisation, too much explanation and questioning by the teacher, and a limited or unsuitable choice of poems to read.

The interest shown in non-fiction, especially by boys, is explained by the number of hobbies and interests which send children to books at this age. Conversation with groups of children showed that some of them found books helpful in connexion with their hobbies as follows :

Boys : Train spotting.
Car numbers and makes of cars.
Aeroplane spotting and aero-modelling.
Sport—rules : autobiographies : hints on how to play or how to train.
Collecting—coins, stamps, labels.

Girls : Sewing.
Wildflowers—collecting and pressing.

Boys and Girls : Drawing of ships, animals, buildings, etc.

TABLE 8

FACTORS INFLUENCING CHILDREN WHEN CHOOSING A BOOK TO BORROW OR BUY.

	Voted most important		Voted least important	
	Boys	Girls	Boys	Girls
1 Reading some of the book first to sample it	358	385	22	8
2 What the book is about	296	256	14	2
3 Books recommended by teacher or parent	105	74	61	48
4 The name of the author	85	107	111	78
5 Previous knowledge of the story from film or television	43	45	54	42
6 A friend's recommendation	24	25	65	43
7 Plentifully illustrated	38	9	594	679

(*b*) *Choosing a Book.*—The factors have been placed in the order of importance assigned to them by the children, giving equal weight to the favourable and unfavourable ratings made by the children.

There was general agreement between the sexes and social areas, except that children in Residential areas were slightly more resistant to teachers' and parents' recommendations than those elsewhere, perhaps because of their greater reading experience.

Discussion with children elicited the existence of strong feelings against the illustration of stories. It was said that a poor illustration could mislead the reader in his attempts to follow the story or to build up a picture of a scene or character, that pictures of incidents in the plot destroyed the element of suspense or interfered with the free working of the imagination. But it was agreed that a picture on a dust-jacket made a book more attractive and could give a fair indication of the subject-interest of the book, and that in works of non-fiction good illustrations were necessary.

The importance placed by children on first dipping into a book or reading a descriptive note about its subject matter is striking. It shows that most children at this stage are able to make their own selection, given a reasonable amount of time at the library shelves. Many of them have been taught to expect that a good book often starts unpromisingly, e.g., *Children of the New Forest, Tom Brown's Schooldays.*

In discussion, children said they would like to spend more time in listening to other children describing what books they liked. It would appear that recommendations of other children are generally rejected because these are not discussed fully enough.

(*c*) *Children's favourite books in private reading.*—Most children of this age have enough reading experience to make a choice of this nature, though some may find it hard to choose between several books. Only books which received five or more mentions are listed.

A.—BOYS.

Fiction.	Total
Enid Blyton's books :	
Various titles	66
" Fives " series	50
Biggles books—W. E. Johns	49
Treasure Island	31
Coral Island	30
Black Beauty	23
Robinson Crusoe	22
Robin Hood stories	19
Tom Sawyer or Huckleberry Finn	18
Tom Brown's Schooldays	13
William series—Crompton	13
King Arthur and his Knights	12
Cruel Sea—Monsarrat	9
King Solomon's Mines	8
Kidnapped	7
Ransome's books	6
Ivanhoe	6
Jennings series—A. Buckeridge	6
Phantom Patrol—Channel	6
Gaunt of the Pearl Seas Patrol—Downie	5
Non-Fiction.	
An Encyclopedia	18
Ascent of Everest—Hunt	7
Bible Stories	6

B.—GIRLS.

Fiction.	
Enid Blyton's books :	
Various titles	162
" Fives " series	87
Black Beauty	57
Little Women	45
What Katy Did	27
Heidi	19
Good Wives	11
Jane Eyre	9
Angela Brazil's books	9
Arthur Ransome's books	8
Uncle Tom's Cabin	8
Heidi grows up	7
Coral Island	6
Heidi's Children	5
Children of the New Forest	5

A few titles appear on both lists, e.g., the " Fives " and " Sevens " series, and *Shadow the Sheepdog* by E. Blyton, Arthur Ransome's series (though in Residential areas only), and a few children's classics—*Coral Island, Black Beauty* and *Uncle Tom's Cabin.*

It is noteworthy that a few children at this age can derive great enjoyment from a few books written for adults, e.g., *The Cruel Sea, Jane Eyre, Ascent of Everest, Reach for the Sky, Jump for Joy.*

(*d*) *Books children had not liked.*—Many children made no return here, a few writing that they liked all their books. Only books which received five or more mentions are listed.

A.—BOYS.

*Biggles series—Johns	15
*Treasure Island	11
Black Arrow	6
*Tom Sawyer and Huckleberry Finn	6
*Black Beauty	5
Oliver Twist	5
*King Arthur	5
*Robin Hood	5
*Tom Brown's School-days	5
*Uncle Tom's Cabin	5
The Water Babies	5

Non-Fiction.

*Encyclopedia	6

B.—GIRLS.

Fiction.

Alice in Wonderland	21
The Water Babies	15
*What Katy Did	11
Treasure Island	11
Wind in the Willows	10
Grimm's Fairy Tales	7
Gulliver's Travels	6
Christmas Carol	6
Dimsie grows up—D. F. Bruce	6
*Uncle Tom's Cabin	6
*Little Women	5
Robinson Crusoe	5
Tom Sawyer and Huckleberry Finn	5
Abbey Girls series—Oxenham	5

The fact that so many children's classics failed to appeal to some children may indicate that they attempted to read them before they were ready for them. With the present good supply of editions of most children's classics this is bound to happen at times when children receive prizes or gifts. It is certainly not a valid reason for excluding books on this list from class libraries, though the findings seem to justify those who give children a large measure of free choice in their reading by having available a plentiful and varied supply of books.

V.—THE EFFECT OF FILM, RADIO AND TELEVISION ON READING.

The questionnaire asked the child : (*a*) to name a book which he had read after seeing a version of it on the cinema or T.V. screen or after hearing it as a wireless play, (*b*) to state how much time he spent each week in looking at television, and how much in reading.

(*a*) *Frequency with which the subsequent reading of books was mentioned.*

Boys		Girls	Total
377	Films	211	588
63	Radio	54	117
244	Television	517	761

Children, especially boys, frequently went on to read adult books after seeing the film—e.g., *The Cruel Sea, Ivanhoe, The Dam Busters, The Ascent of Everest*—

---

* Also listed as favourite books by other children.

or to read children's classics like *Treasure Island*, *Tom Brown's Schooldays*, *Swiss Family Robinson* and the Robin Hood and Arthurian legends.

Radio made a comparatively small impact despite the programmes devoted to introducing the work of contemporary children's writers.

Television stories made a great appeal to girls, especially the *Heidi* and *Heidi Grows Up* series. Nearly 200 girls read *Heidi* as a result of the T.V. serial and 20 boys did so. In discussion it was found that few children tried to anticipate the next instalment by getting a copy of the book. The book was read after the film or serial because it gave a fuller treatment of the story than did the film. Many of the children were aware when reading the book that the director of the film had taken liberties with the story in order to make a more exciting film. Where the printed story was hard to follow, e.g., *Ivanhoe* or *Children of the New Forest*, some children found that the experience of the film or televised play helped them to follow it more easily.

In the great majority of cases, books which children elected to read after the experience of the story in another medium were books of acknowledged high quality, and in this respect the film, radio and television are valuable allies of the teacher and librarian.

(*b*) *A comparison of the amount of time spent at home on reading and on watching television.*

	Boys	Girls
More time spent in reading	19·6%	26·4%
More time spent in watching television	44·2%	35·8%
About equal time spent in reading and watching television	2·9%	4·4%
No television programmes seen regularly or at all	33·2%	31·7%

Children's estimates of the amounts of time spent are not considered to be reliable enough to show how much time each week is spent on an average in reading or viewing. But it is thought that a child's reply may give a fair indication of whether he spent more time on reading than in viewing. Only amongst girls in Artisan A areas was rather more time spent on reading than in viewing. The highest percentage of children spending more time on viewing was in Housing Estate areas.

In discussion, children maintained that time spent in viewing did not reduce the time for reading so much as the time for playing, or in a few cases, the time in listening to wireless programmes. Since much reading at home is reading in bed, it may, however, be argued that an habitually later bedtime is bound to reduce the amount of time for reading.

ACKNOWLEDGEMENTS.—The writer wishes to acknowledge with gratitude the courtesy and interest shown by teachers in co-operating with him in this investigation, and by officers of local education authorities in making the enquiry possible and in suggesting a suitable sample of schools.

(*Manuscript received, March 9th, 1956*).

THE PROSPECT OF MARRIAGE: A STUDY OF THE ATTITUDES TOWARDS FURTHER EDUCATION OF A SAMPLE GROUP OF SECONDARY TECHNICAL AND SECONDARY GRAMMAR SCHOOL LEAVERS

By M. GALLAGHER

(Department of Social Science, University of Liverpool).

SUMMARY.—Many secondary grammar and secondary technical school girls fail to complete the full school course. This enquiry among 190 of such girls indicates how far their expectation of marriage influenced them to leave school early and to enter occupations not requiring further training and study.

I.—INTRODUCTION.

THE Central Advisory Council for Education[1] recently listed the opinion that " a good education is wasted on a girl because she will get married," as one reason why grammar school girls leave school before they have completed the full course. According to Newsom,[2] " Since their period of employment will be temporary and economic security will be provided by the future husband, why spend time and money giving girls a prolonged professional training? The answer that it will make her a better wife and mother or that she may combine marriage and a profession is, unfortunately, not based on scientific evidence."

Nevertheless, grammar schools purport to teach, according to the Norwood Report, " the mind interested in learning for its own sake."[3] In order, therefore, to ascertain how far the prospect of marriage does deter secondary grammar and secondary technical school girls from continuing their education at school and afterwards, a one in four random sample group of such girls was selected from all the 869 girl leavers in the County Borough of Wallasey, Cheshire, during the period July, 1947 to July, 1950. Of these 217, 190 (hereinafter referred to as ' the sample ') were traced and persuaded to co-operate. 163 were interviewed in their own homes and the remaining twenty-seven completed a postal questionnaire. The sample had left school between three and six years before the time of interview.

Wallasey is a residential district in the industrial and commercial area of Merseyside. All kinds of educational facilities and most occupations are available. At the time the girls were interviewed, the local education authority supplied grants to all young people who proved that they were capable of benefiting from further education.

The following questions were asked of the sample. Their answers were classified after all the sample had been interviewed. No answers were suggested to them.

---

[1] *Early Leaving,* Report of the Central Advisory Council for Education (England), 1954.

[2] JOHN NEWSOM (1948). *The Education of Girls,* p. 31. London.

[3] *Curriculum and Examinations in Secondary Schools.* Report of the Committee of the Secondary Schools Examinations Council, 1943, p. 2.

TABLE 1

Questions and Responses of the Sample.

	No.	%
What is your father's occupation ? (Classification according to that used in the 1951 Census.)		
Professional	14	7·5
Intermediate	55	29·0
Skilled	99	52·0
Semi-skilled and Unskilled	22	11·5
	190	100
What academic standard did you attain at school ?		
Higher School Certificate or School Certificate of University Entrance Standard	46	24
School Certificate of lower standard	108	57
Failed or did not enter School Certificate	36	19
	190	100
At what period in the school course did you leave ?		
Completed whole school course	49	26
Did not complete school course	141	74
	190	100
Nos. capable (according to school attainments) of entering VI Form year and who did not do so.	27	14
Why did you not complete the School Course ?		
Dislike of some aspect of School	59	42
Desire for adult status	67	47
Earnings wanted at home	11	8
Services wanted at home	4	3
	141	100
When did your parents want you to leave ?		
Later than child actually did	9	5
Earlier than child actually did	8	4
Let child decide herself	173	91
	190	100
What were your vocational ambitions on leaving school ?		
To enter occupation requiring training and study	156	82
To enter specific occupation (other than above)	13	7
No vocational ambitions	21	11
	190	100
What are your your present vocational ambitions ?		
To undertake an occupation requiring training and study	2	1
To gain promotion in, or make a success of present job	125	66
No particular vocational ambitions	63	33
	190	100

If you have any vocational ambitions, what have you done to realise them ?

	No.	%
Made constructive enquiries and/or undertaken courses other than minimum ones required by present occupation	42	22
Nothing	148	78
	190	100

What is your present occupation ?[4]

One which requires training and study	78	41
Office work	97	51
Occupation other than above	15	8
	190	100

What is your marital status ?

	No.	%
Single	110	58
Married or engaged to be married	78	41
No wish to be married ever	2	1
	190	100

Latest desired age of marriage given by any of the sample = 28.

Would you like to combine your occupation with marriage ?

Yes	49	26
No, not practicable to combine the responsibilities of both	78	41
No, children suffer from lack of mother's care	63	33
	190	100

What is the value of vocational education for a woman ?

None	65	34
More interesting job results	63	33
Economic advantages	30	16
All education intrinsically worth-while	32	17
	190	100
No. of those capable according to school attainments of entering a profession and who did not do so. (This does not include those debarred by family or financial circumstances)	34	18

[4] Following is the list of occupations held at the time of interview. Those who had not completed their training are listed according to their ultimate career.

Occupations Requiring Training and Study.	
College-trained teacher	18
Nurse	21
Laboratory Assistant	20
Graduate Teacher	7
Occupational Therapist	2
Radiographer	2
Doctor	1 each
Physiotherapist	
Shorthand Teacher	
Dancing teacher	
Retail buyer	
Florist	
Actress	
Farmer	
Total	78

Office Work	97
Other Occupations :	
Shop Assistant	10
Children's Nurse	3
Warehouse Assistant	1
Never worked	1
Grand Total	190

From the foregoing information it can be seen that most of the sample expected a fairly short working life. Nearly half were married or engaged to be married between 19 and 24 years and of the remainder all except two hoped to be married before the age of 28 years.[5] 74 per cent. of the sample intended to relinquish their career on marriage.[6] It is interesting that this intention was expressed irrespective of the degree of enthusiasm for the job. Table 2 shows that approximately equal proportions in each type of occupation intended to give up their careers on marriage, despite the fact that the majority of the latter group who had occupations requiring training and study were ambitions, compared with less than half of those in office and other occupations.

TABLE 2

OCCUPATIONS OF SAMPLE WITH REASONS FOR DISCONTINUATION OF WORK AT MARRIAGE AND EXTENT OF VOCATIONAL AMBITION.

	Occupations requiring training and study	Office and other Occupations
Would give up work for sake of children	49%	42%
Difficulty of combining responsibilities	51%	58%
	100	100
	N 57	N 84
Those in above categories ambitious to get on or make success of job	83%	38%
	N 47	N 33

This expectation of a brief working life had several results. Firstly, more than a third thought that further education was of no value for a woman. " You might as well have a good time while you are young," " Studying is a waste of time and money," were typical remarks by no means confined to those of lesser academic ability. Secondly, vocational ambition was lacking in many of the sample. Only two were interested in undertaking a job different from their present one and which entailed greater responsibility. Although 66 per cent. said that they were eager to gain promotion, in or make a success of, their present job only 22 per cent. had done anything constructive about their, ambition. Furthermore, 33 per cent. had no vocational ambitions at all at the time of interview compared with 11 per cent. who had none when they left school. As a personnel manager of a concern employing over a thousand girls of this educational standard said to the interviewer, " They arrive from school full of eagerness and willingness to learn. By the time they are nineteen or twenty they think of nothing but boys and marriage."

Judging from their comments many of the sample were deterred from entering an occupation requiring training and study because of theirexpectation of a short working life. Yet a high proportion of those engaged in office and other occupations expressed dissatisfaction with them compared with those engaged in the more creative occupations ; but as a clerk with a very high school scholastic record said : " Of course, there is nothing to my job, but what is the use of bothering when I shall probably get married ? "

---

[5] This marital condition is reflected in that of the country as a whole, c.f. *The Annual Abstract of Statistics* for 1953, p. 2.

[6] This intention is similar to the occupational status of women in England and Wales. The 1 per cent. tables (Part 1, p. 56) for the 1951 census show that nearly four times as many married women are unoccupied as compared with those occupied; that among professional women nearly three times as many are single as married. Furthermore, a national survey of attitudes of women to employment after the war (G. Thomas, *Women at Work*) revealed that 84 per cent. of the single women would not work after marriage and the reason given most often was the welfare of the children. Cf. also P.E.P., *Graduate Wives*, 1954.

Lastly, many of the sample appeared to have left school early because of their parents expectation that they would get married. Very few of the parents appeared to have encouraged their child to continue with schooling ; yet only in a few cases were the child's earnings or services required. The reasons why the sample themselves wanted to leave school early appear to be more passing symptoms of adolescence than well founded objections, in the view of the Early Leaving Report.[7] Their view is confirmed by the fact that among the Wallasey sample equal proportions from each type of school expressed dislike of school although each school is very different in routine and activities.

II.—CONCLUSIONS.

The information given by the Wallasey sample revealed that 74 per cent. left school earlier than the sixth form ; that less than half undertook occupations which embodied further education ; that of the 67 per cent. who were vocationally ambitious only 22 per cent. had undertaken constructive measures to further their ambitions ; that vocational ambition declined after leaving school ; that three-quarters intended to give up their work on marriage ; that many of these were vocationally ambitious ; that nearly half were married or engaged in the six years after leaving school ; that nearly all the remainder hoped to be married before the age of 28 years ; that more than a third thought vocational education of no value for a woman although many of those in the occupations which embodied no further education were dissatisfied with them compared with very few of those in the more creative occupations ; that the sample considered that the incentives to undertake vocational education were the interest of the ultimate job, the economic security education provided and the intrinsic value of education.

The prospect of marriage appears to hinder many girls from secondary grammar and secondary technical schools from undertaking further education. The abilities of the more intelligent young women in the community are as a result not fully developed and an expensive form of secondary education is often largely wasted. Further research is needed into the question of how best the adolescent girl's interest in further education can be encouraged with particular emphasis on the differentiation of the curriculum for boys and girls.

(*Manuscript received, May 18th, 1956*).

---

[7] *Early Leaving*, *op. cit.*, p. 38, which says that " petty irritation with school is more common amongst girls than boys."

DIFFERENCES BETWEEN SCHOOLS FOR THE MALADJUSTED AND ORDINARY BOARDING SCHOOLS

By M. L. KELLMER PRINGLE*
(*University of Birmingham, Remedial Education Centre.*)

SUMMARY. 1.—A comparative study is presented of some of the physical, administrative and educational conditions prevailing in residential schools for the maladjusted and ordinary boarding schools.

2.—General features, such as size, staffing, the children's age and length of stay are compared and contrasted. The incidence of educational retardation among maladjusted children was found to be much more extensive and the degree of it more serious.

3.—While the incidence of problem behaviour was, of course, much greater in schools for the maladjusted, all the symptoms listed were also found among the pupils in ordinary boarding schools.

4.—All schools for the maladjusted considered child-adult relationships to be of great importance, as well as the maintenance of close liason with the children's homes. More liberal staffing conditions seem essential if these two conditions are to be met satisfactorily.

I.—INTRODUCTION.

IN the Education Act of 1944, where eleven categories of ' handicapped children ' are enumerated, ' the maladjusted ' are specified as a separate class for the first time in any statutory enactment. The act itself does not explain how the term is used, but according to the Ministry's regulation† maladjusted children " require special educational readjustment." Treatment may be given in child guidance clinics, school psychological services, or hospital out-patient units, while the child remains in his own home and continues to attend a day school. A few local education authorities have set up special day schools for maladjusted children. Where a child is too maladjusted to benefit from the abovementioned facilities or where none exist within his locality or where, for any other reason, it is considered essential for the child's readjustment that he be sent away from home for a period, a residential school for maladjusted children or a hostel is recommended. There is, however, a serious shortage of available places particularly for children of secondary school age and those of less than average intelligence. Some boarding schools catering for normal children are prepared to accept a few who are maladjusted.

II.—PURPOSE OF THE INVESTIGATION.

Although the first residential school for maladjusted children was started more than twenty-five years ago, little has so far been published in this country about or by such schools.‡ Therefore, the Research Committee of the Association of Workers for Maladjusted Children decided to carry out an investigation which would provide information about various aspects of residential schools and hostels for maladjusted children. Without adequate information it is difficult to plan future lines of development, and therefore, the aim of this study was mainly to be fact-finding. It was felt that the results would perhaps be more meaningful if a comparison could be made with the conditions, practices and views prevailing in ordinary boarding schools.*

---

* A report of this investigation was prepared on behalf of the Research Sub-Committee of the Association of Workers for Maladjusted Children by M.L.K.P. The following were members of the Research Sub-Committee : P. Dockar-Drysdale, Dr. Portia Holman, M. Horovitch, W. F. Kemp, Dr. M. L. Kellmer Pringle (Chairman), and Otto L. Shaw. This paper is a summary of the full report, copies of which can be obtained from the author.

† *Handicapped Pupils and School Health Service Regulations*. Min. of Education, 1945, H.M. Stationery Office.

‡ However, available publications are listed in the Bibliography. The *Report of the Committee on Maladjusted Children* (H.M.S.O., 1955) was published shortly after this paper was written.

A questionnaire was drafted by the Research Committee, sent to head teachers of schools for maladjusted children for comments and then re-drafted to include their suggestions. The final version, together with an explanatory letter, was sent to forty-nine schools and hostels for the maladjusted and forty-two ordinary boarding schools. Replies were received from thirty-one schools and five hostels for maladjusted children and twenty-seven ordinary boarding schools. With a few exceptions all the schools and hostels belonging to the A.W.M.C. participated, while some schools who have not joined the association, also took part. All the schools for the maladjusted, except one, are residential.

III.—COMPARISON OF GENERAL FEATURES.

The thirty-six schools and hostels for the maladjusted are widely situated geographically and the same is true of the twenty-seven ordinary schools. Both types of school are mainly co-educational and non-sectarian. A greater proportion of the ordinary schools are recognised by the Ministry of Education, while all of them are independent in contrast with those for the maladjusted of which one-third are run by an L.E.A. A greater number of co-educational schools for the maladjusted cater only for children under the age of 11 years as compared with ordinary boarding schools, and the latter are willing and able to admit children both at an earlier and later age.

Schools for the maladjusted are small, catering for twenty to forty-five in the great majority of cases (twenty-eight schools) with only three schools having up to 100 children ; in the ordinary schools numbers are much larger, going up to 400 children, with sixteen schools catering for over 150 and only three schools for less than fifty children. Thus, though a greater number of schools for the maladjusted answered the questionnaire, there are more than three times as many children in the ordinary boarding schools. The total number of pupils in the ordinary schools is 4,686, in the schools for the maladjusted 1,455.†

Both types of schools have waiting lists ; however, as ordinary boarding schools are much larger and there are a greater number available in the country as a whole, the problem is more serious for maladjusted children for whom available places are severely limited. There is a marked difference in the average length of time that children spend in ordinary schools and those for the maladjusted. The great majority of children spend many years, usually during the secondary stage, at ordinary boarding schools. Since schools for maladjusted children aim to return their pupils to their own homes, the more successful the school or the less disturbed the child, the shorter the stay need be. The average length of stay was found to be 2½ years, with a range from 6 months to 6 years.

---

* The description " ordinary boarding schools," was quite justifiably objected to by some head teachers. This term is used as a kind of shorthand to differentiate such boarding schools from schools taking only maladjusted children. It does not imply any value judgment, i.e., that the schools are ordinary in the sense of having no individuality or nothing of interest to offer to their pupils, nor does it imply that all their children are ordinary or normal in the sense of never showing any transient behaviour difficulties.

† In the *Report of the Committee on Maladjusted Children*, the total number of such children maintained by local education authorities (both in independent schools for the maladjusted and local education authorities' boarding special schools) was 2,234, so that this survey covers over half the number of children.

It was decided to compare the cost of wear and tear in ordinary schools with those for the maladjusted. Differential expenditure seems to be most marked with regard to the repair and renewal of furniture in the two kinds of schools, but at the same time differences between individual schools are very considerable (Table 1).

TABLE 1

RANGE OF EXPENDITURE PER CHILD PER ANNUM.

	Maladjusted N=29*	Ordinary N=19*
Premises	£4—£30	£1 5s.—£25
Furniture	£1—£12	4s.—£5 10s.
Equipment	£1 10s.—£18	£1 10s.—£12
Clothes	£2—£24	—**

* Not all schools replied to these items.
** Only four schools replied to this item.

IV.—COMPARISON REGARDING STAFFING.

There is clear evidence that the ratio of staffing is much more generous for maladjusted children, particularly if child care, clerical and domestic staff are included. Thus, in over two-thirds of the schools, there are no more than 2½ children for each staff member, while in schools catering predominantly for normal children the ratio in almost two-thirds of the schools is five or more children per adult. The ratio of teaching staff per child is also more favourable in schools for the maladjusted although this difference is not so marked. There are considerable variations in the duties of staff and there is a general trend for teachers of the maladjusted to have longer teaching hours, more extraneous duty periods and shorter holidays.

V.—EDUCATIONAL AND DISCIPLINARY ASPECTS.

Marked differences emerge regarding the children's educational attainments. Almost all schools for the maladjusted have the results of objective tests of intelligence and attainment made available to them and in many cases carry out testing themselves. On the other hand, most ordinary boarding schools based their answers to the questions regarding educational achievement on teachers' estimates. Schools for the maladjusted state that rarely or never do their pupils work up to capacity, while the position is reversed in ordinary boarding schools. Moreover, the degree of retardation reported is far less serious in the latter : only five schools state that some of their pupils are retarded by at most 1½ to 2 years, while eighteen schools for the maladjusted report having children who are retarded to a more severe degree than this (Tables 2 and 3).

TABLE 2

Work up to capacity	None	rarely	mostly	yes	not known
* Maladjusted Schools (N=31)	21	8	—	1	1
Ordinary Schools (N=27)	5	1	5	15	1

* Hostels excluded.

TABLE 3

Retardation in years	None	Under 1 year	1½-2 yrs.	2 years	3 years or more	Varying degree
* Maladjusted (N=31) ..	—	2	3	14	4	8
Ordinary (N=27)	15	5	5	—	—	2

* Hostels excluded.

All schools were willing to describe quite fully the methods of discipline used. Clearly this is once again a question which cannot be dealt with adequately in a questionnaire. However, to provide a general impression the replies have been classified under a number of sub-divisions and headings (Tables 4 and 5). Not only did all the schools reply to this question but most mentioned more than one of the categories listed. The importance of considering the needs of the individual child is mentioned more often by schools for the maladjusted, as one would expect from the function of such schools. Relatively small numbers and small classes also make it more practicable " to fit the punishment to the criminal " than is possible in a large boarding school. For similar reasons, schools for the maladjusted use more frequently or exclusively such means of discipline as are employed by ' good families,' e.g., relying on the personal relationships between children and staff, discussing problem behaviour with the children themselves and where possible, preventing difficulties from arising.

TABLE 4

THERAPEUTIC METHODS OF DISCIPLINE.

	Self-government	Self-discipline	Suited to individual	" Good family " methods	Restitution	Relation between child and staff	Anticipation and Prevention	Discussion of problems
Maladjusted (N=36)	3	3	11	5	3	9	3	3
Ordinary .. (N=27)	4	4	5	2	4	5	—	—

TABLE 5

RETRIBUTIVE METHODS OF DISCIPLINE.

	Loss of privileges	Loss of sweets or puddings	Sent to bed early	Fines	Fatigues	Detention	Corporal punishment ' rarely '
Maladjusted ...... (N=36)	11	5	2	4	3	—	9
Ordinary .... (N=27)	14	3	3	5	5	4	2

VI.—INCIDENCE OF PROBLEM BEHAVIOUR.*

While of undoubted interest, it was considered impracticable to obtain a complete clinical picture of each child in the schools for the maladjusted. Filling in a lengthy questionnaire was in any case likely to be a burden to busy head teachers. To include an exhaustive list of possible symptoms would have made matters worse. Instead, seven broad categories were chosen which, in the view of the research committee, represented the more serious symptoms of childhood neurosis. Inevitably, such a short list leaves out more than it includes and omissions readily come to mind. The following categories were chosen: abnormal sexual behaviour; encopresis; enuresis; asthma; withdrawn and inhibited; depressive symptoms; stealing. The number of schools who reported that any of these symptoms are exhibited by their pupils is shown in Table 6.

TABLE 6

NUMBER OF SCHOOLS MENTIONING THE OCCURRENCE OF THE SYMPTOMS.

	Enuresis	Stealing	Withdrawn	Depressed	Encopresis	Abnormal Sex Behaviour	Asthma
Maladjusted (N=36)	36	32	31	29	26	22	12
Ordinary .. (N=27).	18	20	16	8	4	7	19

Enuresis is the most prevalent problem in schools for the maladjusted, being listed by all the schools and hostels. Asthma is mentioned by the smallest number, which may partly be due to the fact that open-air schools are an alternative placement for asthmatic children. On the other hand a considerable number of ordinary boarding schools have children suffering from it and, except for stealing, asthma is mentioned by the greatest number. (It is interesting that many ordinary boarding schools substituted 'petty pilfering' for stealing.) Most schools mentioned more than one of the above problems. However, when the actual number of children showing these symptoms is considered, it can be seen (Table 7) that in both kinds of schools stealing and enuresis are the most common problems. In fact, the order of frequency with which these symptoms are shown by the children in the two kinds of schools is the same for all the symptoms except asthma, which takes fourth place in ordinary schools instead of depression.

The comparative incidence of these symptoms is, of course, very different in the two kinds of schools. The percentage for all the symptoms is about 7 per cent. in ordinary boarding schools and about 60 per cent. in schools for the maladjusted.† There are, also very considerable variations from school to school. One might have expected a higher incidence of problems in maladjusted

---

* In a covering letter it was suggested to ordinary boarding schools that most of this section was probably not applicable to them; yet the great majority of schools took the trouble to reply to these questions.

† Many of the ordinary boarding schools state explicitly that the proportion of maladjusted children is limited to 10 per cent. or 15 per cent. of the total school population to maintain a workable balance and to enable the 'difficult' child to be absorbed smoothly into the school life.

TABLE 7

NUMBER OF CHILDREN EXHIBITING SYMPTOMS.

	Total No. of children	Total No. of Problems	Stealing	Enuresis	Withdrawn	Depressed	Encopresis	Abnormal Sex Behaviour	Asthma
M.	1455	900	248	222	163	123	60	57	27
O.	4686	295	94	76	58	16	7	9	35

schools (Table 7) and a higher overall percentage. It must be remembered, however, that only a limited number of symptoms was asked for, so that many major and frequent problems were excluded (for instance stammering, nightmares, food fads, aggressiveness, violence, tics, etc.). Moreover, a school for maladjusted children contains at any one time some new entrants who, presumably, will be among the most seriously disturbed, a proportion of children who have improved considerably owing to treatment, and some who are near discharge.

VII.—DISCUSSION OF MAIN FINDINGS.

(*a*) *Residential Provision for Maladjusted Children.*—Perhaps the feature most common to schools catering for maladjusted children is the emphasis placed on personal relationships between child and adult. It has been stated that " the cause of maladjustment is not so much the lack or loss of a parent or parents as the lack or loss of a relationship with a parent. The result is that the child himself has his capacity to form relationships interfered with . . . The function of a school (for maladjusted children) is to restore these relationships, but in the most severe cases the capacity for making relationships has to be developed or restored before the relationships themselves can be improved."* Should this be true (and recent research findings by Bowlby and Spitz substantiate this view), then it is essential that schools for the maladjusted should continue to remain relatively small units as is the case at present. Thus, while there seems to be a need for considerably more places than are available for the residential placement of maladjusted children, it would be unwise to provide them by enlarging present schools. Though it is not possible from the evidence obtained to judge the exact extent of the shortage of places, there is little doubt that it is serious. A number of schools do not even keep a waiting list, since applications greatly exceed possible vacancies ; it would be nonsensical for a school to keep a list which might well contain a greater number of children waiting than on roll at present, since several years would elapse before sufficient vacancies would occur.

(*b*) *Child-Adult Relationships.*—The importance given to child-adult relationships is indicated by the fact that, although the child/staff ratio is much more favourable in schools for maladjusted children, their staff have considerably less free time during the week and shorter holidays throughout the year. This is partly due to the fact that maladjusted children are less capable of being left to their own devices, and partly that many are not able to go home for the whole of the holidays because of the severity of their maladjustment. At the same time the day-to-day work entails more continuous and thus closer child-adult contacts. In view of the considerable emotional strain imposed by the work—maladjusted children making much greater demands on the tolerance and stamina of the staff by the very nature of their handicap—it seems desirable for the sake of the mental health of the staff that their conditions of work be as favourable, if not more so, than those of their colleagues in ordinary boarding schools.

---

* Unpublished report submitted by the A.W.M.C. to the Ministry of Education Committee of Inquiry on Maladjusted Children.

The close relationship between emotional difficulties and educational underfunctioning is shown by the fact that most schools for the maladjusted find their pupils to be seriously retarded in the basic subjects. There is no clear-cut line of demarcation between emotional and educational difficulties but on the contrary a close inter-relationship (for a discussion see Kellmer Pringle, 1954).

Most schools recognise this problem and have developed special schemes or features to enlist their pupils' co-operation in the educational sphere.

(*c*) *Liaison with the Home.*—Since the main aim of schools for maladjusted children is to return their pupils to live with their own families, it must be regarded as a serious weakness that shortage of staff prevents as close a liaison between parents, clinics and schools as would be desirable for full and speedy recovery. The appointment of a case worker, if necessary shared by two or more schools, to be responsible for contact with and treatment of the home while the child is away would seem at least as urgent a problem as the provision of more places. This would be a natural development growing out of the work at present being done by P.S.Ws. in child guidance clinics. It would in no way curtail the contribution made by them, but on the contrary strengthen the links and make possible more frequent contacts between schools and clinics on the one hand, and schools and the children's parents on the other hand.

(*d*) *Discipline and Punishment.*—Many schools for the maladjusted also emphasise the importance of personal relationships in connection with discipline and methods of punishment. Some of the findings, however, are perhaps rather unexpected. Though many schools explicitly state that corporal punishment is never used, nine schools for the maladjusted resort to it, although very rarely, compared with only two ordinary boarding schools. Three observations seem relevant: firstly, maladjusted children because of their very maladjustment are much more likely to provoke or challenge the endurance of the staff; moreover hysterical outbursts may, on occasions, make a slap essential. Secondly, an analysis of the nine schools for the maladjusted which do occasionally use corporal punishment shows that six are L.E.A. schools (out of a total of thirteen), and three are independent schools (out of a total of twenty-three). Thus, in nearly half the L.E.A. schools this method of punishment is employed (though infrequently) and it might be worth investigating why the incidence is so much higher in these schools than in the independent schools. In the view of the A.W.M.C. there is no place for corporal punishment in schools for maladjusted children. Thirdly, the low incidence of corporal punishment in the ordinary boarding schools, together with the very enlightened methods which are used, are likely to be due to the fact that this is not a random sample of independent boarding schools in this country. On the contrary, the schools which were asked to co-operate in this research project are among the most progressive and liberal in outlook and those which actually took part are perhaps the cream of these. In contrast, practically all schools for the maladjusted in this country were approached and the majority consented to co-operate, so that they can be regarded as a very representative as against a highly selected sample.

VIII.—References.

Bloom, L. (1955). Some aspects of the residential psychotherapy of maladjusted or delinquent children. *Brit. J. Delinquency*, 6,41-49.

Bowlby, J. (1946). *Forty-four Juvenile Thieves.* London : Bailliere.

Bowlby, J. (1951). Maternal care and mental health. *Bull. World Hlth. Org.*, 3, 355-533.

Burn, M. (1956). *Mr. Lyward's Answer.* London : Hamish Hamilton.

Dockar-Drysdale, P. (1953). Some aspects of damage and restitution. *Brit. J. Delinquency*, 4, 4-13.

Pringle, M. L. Kellmer (1954). The remedial education centre. *Brit. J. Phys. Med.*, 17.

Shaw, O. L. (1940). *School Discipline—A Practical Application of Evolved Social Legal Methods.* London : Stockwell.

Shaw, O. L. (1947). *A Brief Study of Re-Education Factors in a Boarding School for Delinquent Children and Adolescents.* London : Cooper.

Special Correspondent (1951). Life in the Caldecott Community. *Times Educ. Suppl.*, Nov. 9th.

Spitz, R. A. (1946). The smiling response. *Genet. Psychol. Monogr.*, 34, 57-125.

Wills, D. W. (1941). *The Hawkspur Experiment.* London : Allen and Unwin.

Wills, D. W. (1947). *The Barns Experiment.* London : Allen and Unwin.

(*Manuscript received, September 17th, 1956*).

SYMPOSIUM: THE USE OF ESSAYS IN SELECTION AT 11+

IV.—EXPERIMENTAL INQUIRIES INTO THE USE OF ESSAY-TYPE ENGLISH PAPERS.*

By D. A. PIDGEON and ALFRED YATES
(*National Foundation for Educational Research in England and Wales.*)

SUMMARY. 1.—A number of essay-type English examination papers were administered on different occasions to groups of eleven-year-old children. Seven examiners were employed in marking the scripts, and a complex experimental design enabled a variety of statistical analyses to be carried out.

2.—In another area the validity of similar papers was investigated by comparing the results with a criterion of success in secondary schools after the children concerned had completed two years of their course. The correlations obtained were also compared with those between an objective test of English and the criterion.

3.—The results of the reliability investigation indicated that:

(*a*) There are differences between examiners with respect to the mean mark and the variance of marks awarded for English papers of the essay type.
(*b*) The seven papers given varied in their level of difficulty, as judged by the mean mark awarded.
(*c*) English papers of the type described are subject to practice effects.
(*d*) There is clear evidence from this experiment that examiners and children have idiosyncrasies which interact.
(*e*) The error variance attributable to examiners is greatly reduced on the more objective English question paper (Part II) compared with the essay (Part I), although the amount of fluctuation due only to the children is relatively of almost the same size.
(*f*) There were significant differences between examiners regarding their *consistency* of marking.
(*g*) Compared with the figures of ·980 for the strictly objective Moray House English tests and ·970 for the slightly less objective National Foundation English tests, the reliability of the 'old type' English paper considered here was found to be approximately 0·88.
(*h*) The equivalent test, re-test reliabilities for the essay alone were ·766 if the *same* examiner re-marked, and ·719 if a *different* examiner re-marked.
(*j*) The upper limit to the reliability of the 'older type' of English paper, including an essay, is set by the figure of 0·923 calculated on the assumption of perfect marking.
(*k*) Slightly over half the children opted to write on a different essay topic when attempting the same paper a second time. An analysis of the children's choices, however, showed that for some of the papers their choice was severely limited by the inclusion of one or more unpopular topics.

4.—The results of the validity investigation indicated that the 'old type' English examination provides a significantly less satisfactory forecast of subsequent success in secondary schools than is obtained from the results of an objective test.

5.—The introduction of new types of item into objective tests is suggested as a means of creating a less undesirable 'backwash' on the primary school curriculum. Tests of this kind, it is claimed, could maintain the high reliability and validity of the more conventional objective tests.

I.—INTRODUCTION.

THE inquiries described in this article were prompted by some of the criticisms that have been levelled by teachers, education officers and others against objective tests of attainment in English of the type commonly used in eleven plus examinations. Those who bring forward these criticisms do not, in the main, question the reliability or the predictive efficiency of objective tests. Their strictures apply, rather, to the narrowness of the range of linguistic skills that such tests measure, and to the restrictive influences that they, therefore, exert on the primary school curriculum. There is evidence to show that, in some primary schools in areas where the examining authorities rely exclusively on objective tests, little or no opportunity is afforded to the pupils, during the last two years of their course, for expressing themselves in written consecutive English. This had led to the need in many secondary schools, including grammar schools,·to offer what may justifiably be described as remedial teaching to some of even their ablest entrants.

---

* One of the inquiries described in this article was reported in a paper read before the Annual Conference of the British Psychological Society held at Manchester in April, 1956.

One consequence of this dissatisfaction is the tendency on the part of an increasing number of local education authorities, who have for years used objective tests in their selection procedures, to re-introduce the older type of paper that allows for a certain amount of constructive response and consecutive writing and which involves some degree of subjectivity in its assessment. Other authorities, encouraged by the work of Wiseman (1949a and b) and others have sought to mitigate the unfortunate backwash effect of objective tests by adding an English essay to their procedures, but have refrained from reverting entirely to the older form of examination out of respect for the memory of Hartog, Rhodes and Burt (1936) and Ballard (1935).

In view of these trends it was decided by the National Foundation to conduct a series of experiments designed to yield evidence concerning the reliability and validity of the kind of English examination that nearly half of the local education authorities in England and Wales have either retained or restored, and to relate the results to those obtained from other relevant investigations.

II.—AN INVESTIGATION INTO THE RELIABILITY OF ESSAY-TYPE ENGLISH PAPERS.

(i) *Design and conduct of the investigation.*

This investigation involved the administration of a series of English papers, used in recent years by an authority in the South of England, to a sample of some 350 children from a neighbouring authority, who were, for the purposes of the experiment, assigned within schools cyclically to seven groups from alphabetical lists. Each paper consisted of two parts : Part I was an essay on one of three or four proffered topics ; Part II consisted of questions to test the comprehension of selected prose passages, exercises in English usage, punctuation, *et cetera.* All the items were of the creative response type. Some of them admitted of such a restricted range of correct answers as to allow of a nearly objective form of scoring. Others permitted some latitude to the examainer in making his assessments.

The children were all drawn from the final year group of a number of primary schools selected so as to form a reasonably representative sample in terms of range of ability and socio-economic background. The examiners, seven in number, five men and two women, had all had considerable experience of teaching children in the age group concerned. Five were heads of primary schools ; one was a secondary modern school head who had previously taught in primary schools for many years ; one was a grammar school master responsible for teaching English to the lower forms.

The details of the experimental design * are shown in Table 1. The 'occasions' shown in this table are occasions of *marking.* The capital letters

---

* This design was suggested by the Foundation's statistical consultant, Dr. D. J. Finney. The subsequent analyses have been carried out by the Foundation's statistical officer, Dr. P. M. Grundy.

A . . . G indicate the seven papers used in the experiment, the small letters denote the examiners. Each group of children were set three different papers at weekly intervals and were then re-examined with the paper that had been administered to them originally. It will be seen that arrangements were made for the same pair of examiners to mark both administrations of the first paper set to each group. This same pair of examiners also re-marked the scripts that they had marked on the first occasion. The design is complex and it is not possible within the limits of a short report to enumerate all its features. The results dealt with are those that relate to the consistency of the examiners (both inter-individual and intra-individual consistencies), the effects of practice on the successive performances of the groups of children concerned, and the reliability of the examination papers.

TABLE 1

DESIGN OF THE EXERIMENT.

Group	Occasion				
	I	II	III	IV	V
1	Aae	Bac	Dab	Aae	Aae
2	Bbf	Cbd	Ebc	Bbf	Bbf
3	Ccg	Dce	Ecd	Ccg	Ccg
4	Dda	Edf	Gde	Dda	Dda
5	Eeb	Feg	Aef	Eeb	Eeb
6	Ffc	Gfa	Bfg	Ffc	Ffc
7	Ggd	Agb	Cga	Ggd	Ggd

Occasions : IV= re-setting papers of Occasion I
V=re-marking scripts of Occasion I
Papers : A, B, C . . . . G
Examiners : a, b, c . . . . . . g

(ii) *Results.*

(*a*) *Analyses of Means and Variances.*—The treatment of the data falls naturally into two main sections : one dealing with the means and variances of the marks ; the other dealing with the analysis of variance and the different types of correlations obtained. Throughout the analysis, children for whom some of the results were missing were omitted, and the results are, therefore, based on the scripts of 254 children only.

First, we shall consider the evidence relating to differences between examiners derived from an analysis of the mean marks and the variance of marks. The experiment involves seventy combinations, including repetitions, of the four factors—examiners, occasions, groups of children, and papers. For each combination the mean mark and the variance were calculated and the assumption was made that the four factors concerned exert additive effects. The seventy values of the mean and variance fall into pairs belonging to the same group of children, occasion and paper, so that estimates of the differences between examiners can be obtained from the differences within pairs. The adjusted means and variances are given in Table 2 below. It will be seen that, with respect to the means, the examiners fall into a lower group comprising d, b, c, e and a, and an upper group comprising g and f. The differences *within* these groups are not significant, but the difference *between* the groups is highly significant. It is interesting to note that the two examiners f and g, comprising the upper group, were the only two women involved in the experiment. These same two examiners, it will be seen from the table, provided the smallest variance values. The difference between these values for f and g and those for the group comprising b, d and e is significant. There is not, however, in this instance a clear demarcation of all the examiners into two groups as occurred with respect to the means.

TABLE 2

ADJUSTED MEANS OF EXAMINERS.

Examiner	a	b	c	d	e	f	g	All
M	81·37	78·77	79·37	76·40	80·46	89·14	88·73	82·03
V	1303	1422	1254	1448	1454	1013	1090	1284

This same section of the analysis also provides evidence concerning the differences between the examination papers. Since estimates of the differences between examiners are available, we can use these to adjust the seventy values of the mean to what could be expected from a hypothetical average examiner. The adjusted means for the different papers, freed from the effects of differences between examiners, between groups of children and between occasions, are shown in Table 3 below. The difference between paper A, the easiest, and paper E, the most difficult, is significant. These differences are important if the performance of a group of children on one paper is to be compared with that on another without steps being taken to render the marks comparable by, for example, a process of standardisation.

TABLE 3

ADJUSTED MEANS OF PAPERS.

Paper	A	B	C	D	E	F	G
M	89·78	77·26	79·53	81·53	76·98	86·33	82·82
V	1469	1288	1091	1431	1183	1412	1111

Evidence concerning the effects of practice is shown in Table 4. In this table the mean and variance values for the five occasions of marking are shown. In this case no adjustment is necessary because all the examiners, all the groups of children and all the papers occur with equal frequency on each occasion. The differences between occasions with respect to the mean values are significant; those with respect to the variance values are not. The difference between the mean on Occasion I (80·15) and that on Occasion II (84·01) is significant at the 5 per cent. level; the difference between the mean on Occasion I and that on Occasion V (74·99) is significant at the 1 per cent. level.

TABLE 4

MEANS OF OCCASIONS, WITH STANDARD ERROR.

Occasion	I	II	III	IV	V	Standard Error
M	80·15	84·01	84·50	86·51	74·99	1·15
V	1212	1293	1306	1303	1304	64·3

The increase in mean score from Occasion I to Occasion IV resembles the familiar practice effect observed in similar experiments involving objective tests. As is usually found in such experiments, most of the increase takes place between the first occasion and the second. The size of the difference between the first and the final test (Occasions I and IV)—6·36 points—is somewhat less than that usually obtained in similar experiments with objective tests, but nevertheless, demonstrates that papers of this type are affected by practice and presumably by coaching also. The results also indicate that the standards adopted by the examiners were affected by the pupils' improved performances, so that when confronted a second time (Occasion V) with the scripts that they had marked earlier (Occasion I) they betrayed a tendency to apply higher standards and consequently to award lower scores.

(*b*) *Anaylses of Variance.*—We now turn to consider the results of the analyses of variance that were carried out on the data concerning Occasions I, IV and V only. Two separate analyses were carried out, first on the total score, and then on the scores for the two parts of the paper separately. Since for any one group of children the papers on all three of these occasions were the same (see Table 1) the experimental factors affecting a particular child's results were :

the occasion of sitting ;
the occasion of marking ;
and the identity of the examiner.

Not only may there be average effects of these factors but also effects of the two combinations which vary independently in the experiment, that is,

examiner and occasion of sitting ;
and examiner and occasion of marking.

In comparing the marks of a single group of children we have to consider how far these influences vary from one child to another, so that there are six relevant factors and combinations of factors (shown in Table 5).

TABLE 5

STATISTICAL FACTORS CONSIDERED IN ANALYSIS OF VARIANCE.

1. The child.
2. The child and occasion of sitting.
3. The child and occasion of marking.
4. The child and examiner.
5. The child and examiner and occasion of sitting.
6. The child and examiner and occasion of marking.

The provisional assumption made in the analysis of variance of any one group is that, apart from components that are the same for all children in the group, the marks obtained are the sum of independent contributions of these six factors, these being normally distributed about zero means with variances σ_1^2, σ_2^2,...σ_6^2 The different groups were analysed separately and the results pooled. No attempt was made to inter-relate the marks of different groups.*

TABLE 6

COMPONENTS OF VARIANCE AND THEIR STANDARD ERRORS.

Variance	Part I		Part II		Total	
	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.
σ_1^2	265·32	27·4	343·09	33·9	1,132·55	108
σ_2^2	38·13	5·7	55·38	5·2	94·65	11·1
σ_3^2	—2·89	3·2	—0·74	0·40	—2·17	3·8
σ_4^2	17·37	3·6	1·94	0·44	17·03	4·2
σ_5^2	—1·27	4·9	—0·40	0·63	0·98	6·0
σ_6^2	52·38	4·7	6·88	0·62	61·79	5·6
Total ..	369·04	—	406·15	—	1,304·83	—

The main results ,shown in Table 6, indicate that the components of variance for which there is significant evidence are σ_1^2, σ_2^2, σ_4^2, and σ_6^2.

It is useful here to formulate the idea of the *considered opinion* of a particular examiner as the average of a large number of markings of a given script by that examiner. The process of averaging employed eliminates the components of variance associated with σ_3^2 and σ_6^2. Similarly, the *steady performance* of a particular child on a given paper can be thought of as the average of scripts produced by hypothetically re-setting the paper *ad lib.*, and, in this case, the averaging process eliminates σ_2^2 and σ_5^2. When both processes are applied the only part of the error variance remaining is σ_4^2. The significance of σ_4^2 indicates that there are differences between the ' considered opinions ' of examiners regarding the relative ' steady performances ' of different children. In other words, it would seem that the examiners and children have idiosyncrasies which interact. This moreover is the only inter-examiner component of error variance, since σ_5^2 is not significant as it would be if there were any tendency for considered disagreement between examiners about the relative merits of successive scripts produced by the same child.

A noteworthy feature of the results shown in Table 6 is the relative smallness of the last four components of variance for Part II compared with Part I. It is apparent that in Part II the components of error attributable to the examiners are reduced, whilst the amount of fluctuation due only to the children, represented by σ_2^2, is comparable with that in Part I.

It is possible from our data, for any type of variation, to calculate the mean products of Parts I and II in the analysis of variance sense, and thence to derive correlations between the two parts. Thus, the correlation between Parts I and II taken over all the data from Occasions I, IV and V is 0·684. If, however, we average the six marks for each child from these occasions, the correlation between the parts is 0·779.

---

* There is insufficient space in this article to give further details of the statistical analyses. However, the Foundation's statistical officer, Dr. P. M. Grundy, proposes to report elsewhere the points of interest in the statistical methods concerned.

There is also some evidence from this source of a ' halo effect ' operating on examiners. That is, the impression gained by an examiner from marking Part I (the essay) is carried over to the marking of Part II. To overcome this effect, it would be necessary for examiners to mark the two parts of the paper on separate occasions.

(*c*) *Correlations*.—We now come to the various correlations derived from the data and shown in Table 7. In order that comparisons can be made with other work, we have included in this table certain correlations obtained by other investigators, but confined ourselves to those who have dealt with children in the 11+ age group.

For the total paper separate correlations were averaged by transforming to Fisher's z, averaging the z's and transforming back. For the total paper and for the parts separately, the analysis of variance of Occasions I, IV and V were used to reconstruct pooled correlation coefficients whenever possible. Column 8 of Table 7 gives the correlations from both sources, and it will be observed that the agreement is very satisfactory indeed.

The figures for the essay alone of ·814 and ·866 (column 6) for correlations of type (i) and (ii) respectively agree fairly with those reported by previous investigators. Wiseman's (1949b) sample of children were somewhat homogeneous compared with our own and others, and Peaker's figure of ·91 quoted by Vernon and Millican (1954) was, apparently, obtained with very experienced markers. For the total paper, thirty-five separate correlations of type (i), ranging from ·907 to ·971, were obtainable from the data and these averaged to ·945 with no significant difference between them, except for some slight indication of greater inter-examiner consistency with Paper A. Similarly, fourteen correlations of type (ii), ranging from ·893 to ·987, yielded an average of ·964 for the whole paper, and this time the differences were highly significant. Examiner e, a primary school head with many years' experience as an examiner, was the most consistent marker, significantly better than marker b, another primary head with no such experience. As Finlayson (1951) points out, these consistency coefficients are to be compared with unity for the objective test, and presumably the high figure for Part II of ·985 reflects its more objective nature, and thus confirms the results obtained from the analysis of variance.

The differences between examiners e and b regarding consistency are confirmed from correlations of type (v)—the re-marking by the same examiner of one paper set on two occasions (I and IV). Again marker e is significantly better than b.

As explained earlier, one of the main aims of the experiment was to determine the reliability of this type of examination compared with that of an objective test. The calculation of pooled correlations of types (iii) and (iv) for the essay separately have not been carried out, and therefore, we are unable to make direct comparisons with Finlayson's figures (1951). However, it seemed to us that the nearest approach to the test, re-test correlation of objective tests would be given by types (v) and (vi) where the same paper was re-set and re-marked.

Here it will be noticed that so far as the essay alone is concerned the correlation of ·766 obtained when the same examiner remarks the re-set paper drops to ·719 when marked by a different examiner. The corresponding figures for Part II, ·850 and ·845, and for the total paper, ·881 and ·868, are not significantly different.

It is these figures which must be compared with the reliability of objective

TABLE 7

Type of Correlation	Wise-man	Finlay-son	Vernon and Morrison	Peaker	N.F.E.R.			Moray House Objective Test	N.F.E.R. 11+ English Test
(1)	(2)	(3)	(4)	(5)	Part I Essay (6)	Part II Quest-tions (7)	Total (8)	(9)	(10)
(i) *Different* examiners marking the *same* papers	·71	·738	·760	·91	(·815)	(·979)	·945 (·939)	1·00(?)	Probably ·995
(ii) *Same* examiner re-marking the *same* papers	·75 ·928 (pool of 4)	·810 ·939 (pool of 4)	·860	—	(·866)	(·985)	·964 (·954)		
(iii) *Same* examiner marking *different papers*	—	·691 ·863 (pool of 6)	—	·81	—	—	·885	·962	—
(iv) *Different* examiners marking *different* papers	—	·786 (pool of 3)	—	—	—	—	·870		
(v) *Same* examiner re-marking the *same* paper *re-set*	—	—	—	—	(·766)	(·850)	·886 (·881)	·980 (K-R formula)	·970 (K-R formula)
(vi) *Different* examiner marking the *same* paper re-set	—	—	—	—	(·719)	(·845)	·876 (·868)		
(vii) Re-setting the *same* papers and assuming *perfect* marking			—	—	(·874)	(·861)	(·923)		

NOTE: All correlations given are averages, except where stated.
The first figures given in column 8 were obtained by transforming to Fisher's z, averaging the z's and transforming back.
The figures in brackets are pooled correlations, reconstructed from the analyses of variance.

tests, and it will be observed they are considerably lower. The manuals to Moray House English tests quote a number of correlations obtained by giving groups of children two of the tests of their English series, and the average of eleven such correlations was worked out to be ·962 with an average of forty-eight days between the two testings. The average reliability calculated by Kuder-Richardson formula for eight Moray House tests was ·980 and for five National Foundation English tests ·970. It should be explained here that the National Foundation English tests are not strictly objective. Most of the items are of the ' creative response ' pattern, and about 20 per cent. of them require the construction of a sentence or part of a sentence.

One final comment might be made about the correlation of type (vii) listed in Table 7—the test, re-test correlations assuming perfect marking. While these will never be obtainable in actual practice, they do, perhaps, set the upper limit on what could be achieved from this type of paper.

(*d*) *Choice of Essay Topics.*—Although not directly relevant to the main purpose of the investigation, a separate analysis was carried out of the choice of essay topic on the two occasions when the pupils answered the same paper (Occasions I and IV). It will be recalled that for each paper the children were given a choice of three or four topics. The analysis revealed that on the second occasion for all papers 58·5 per cent. of the children wrote on different topics, though for the papers separately this figure varied from 32 per cent. to 70 per cent. Table 8 gives the frequencies of choice of alternative topics for each paper separately, and it is clear from this table that essay topics varied considerably in their appeal to the children. Two of the topics given in Paper B, for example, were, so unpopular that the choice of essay for most of the children taking the paper was limited to one. Clearly, to be fair to all children, topics should be chosen that are of approximately equal popularity.

TABLE 8

FREQUENCIES OF CHOICE OF ALTERNATIVE TOPICS.

Paper	Occasion	Topic			
		(i)	(ii)	(iii)	(iv)
A	First	17	4	27	—
A	Second	20	7	21	—
B	First	37	5	8	—
B	Second	40	8	2	—
C	First	0	11	5	21
C	Second	2	14	8	13
D	First	9	10	22	—
D	Second	11	12	18	—
E	First	3	13	8	19
E	Second	7	13	7	16
F	First	9	24	6	—
F	Second	14	15	10	—
G	First	14	10	8	6
G	Second	15	10	4	9

III.—AN INVESTIGATION INTO THE VALIDITY OF ESSAY-TYPE ENGLISH PAPERS.

Circumstances did not permit us to carry out a follow-up study of the pupils involved in the investigation into the reliability of essay-type English papers. In order, therefore, to obtain evidence concerning the validity of this type of examination, an inquiry was conducted in a county borough in which both objective tests and 'old type' examinations in English and Arithmetic were used in the eleven plus examination. In this area a preliminary selection of candidates is carried out by administering two verbal intelligence tests. The children who take the main examination, therefore, form a sample that is somewhat restricted in range of ability.

(i) *Conduct of the investigation.*

Orders of merit were obtained from the heads of the two grammar and six modern schools in the borough for a total of 473 children who had completed two years of their secondary school course. The orders of merit were scaled and quantified with reference to the average of two preliminary intelligence tests that had been administered as part of the selection procedure two years earlier. The measures of progress so derived were compared with the scores that the children had obtained both in the objective tests and in the old-type examinations that together formed the main part of the selection procedure. The results were submitted to an analysis of variance and both ignoring schools and within schools correlations were obtained.

A full account of this investigation will be reported elsewhere.* The results that we discuss here are those relating to the predictive effectiveness of the essay-type English papers compared with that of an objective English test.

The English examination used in this inquiry was similar in content to that used in the investigation reported above. Part I consisted of questions designed to test the comprehension of passage of prose and exercises in English usage; Part II consisted of an essay.

(ii) *Results.*

Table 9 shows the correlations with the criterion for both parts of the English examination and for the objective English test, and also shows the effect of differences in age between the candidates. The boosted correlations represent an estimate of what the within schools correlations would be if, after allocation, secondary schools had no differential effects on the criterion.

TABLE 9

CORRELATIONS OF ENGLISH EXAMINATIONS WITH THE CRITERION, WITHIN SEX.**,
(*a*) ignoring schools
(*b*) boosted from within schools analysis
(*c*) with age eliminated from (*b*)

Predictor	(*a*)	(*b*)	(*c*)
Age	0·144	0·155	—
Objective English	·724	·671	·685
Essay-type English (I)	·578	·538	·526
Essay-type English (II) ..	·549	·520	·506
English (I)+English (II) ..	·640	·600	·588

---

* *Selection for Grammar Schools*, Third Interim Report on the Allocation of Primary School Leavers to Courses of Secondary Education, by ALFRED YATES and D. A. PIDGEON. Newnes Educational Publishing Co., Ltd. (To be published shortly.)

** The correlations shown in this table have *not* been corrected to offset the effects of the restriction of the sample described above.

The effects of age were found to be significant for the essay-type English examination but not for the objective test. The differences between the correlations with the criterion of the two kinds of test are significant.

IV.—DISCUSSION AND CONCLUSIONS.

The work of McClelland (1942), Emmett (1954), and more recently, Richardson (1956), indicated that the outlook for the older type of English examination was not quite so bleak as the earlier inquiries of Valentine (1932), Hartog and Rhodes (1941) and others tended to suggest. The results of the experiments that we have outlined, however, show that even in ideal conditions, which cannot in practice be contrived,—that is, with a faultless system of marking—papers of this kind do not achieve the level of reliability that is maintained by objective tests, nor do they achieve the same degree of validity.

It would appear, therefore, that authorities are confronted with the difficult choice between adequate instruments that seem to have unfortunate backwash effects and less reliable and valid forms of assessment that may conceivably—although there is no firm evidence on this point—exert a more desirable influence on the education of primary schools.

A close scrutiny both of the kinds of examination paper that were used in these experiments and of others that are currently used by education authorities has convinced us that the difference between the *contents* of such examinations and those of objective tests is by no means so large as has been supposed. The difference is one of format and arises chiefly from the fact that one type of test relies for the most part on creative response items and the other on those of the multiple choice type. One possible solution, therefore, may be to construct tests of attainment in English which maintain a high degree of objectivity in assessment, and consequently an adequate level of reliability, but which contain a large proportion of creative response items. The results of other rseearches that have been carried out at the Foundation suggest that such a compromise is possible and that the reliability and validity of these ' new type ' objective tests is no less than that of the tests of the more conventional objective pattern.*

V.—REFERENCES.

BALLARD, P. B. (1935). *The New Examiner.* University of London Press.

EMMETT, W. G. (1954). Secondary modern and grammar school performance predicted by tests given in primary schools. *Brit. J. Educ. Psychol.*, 24, 91-98.

FINLAYSON, D. S. (1951). The reliability of the marking of essays. *Brit. J. Educ. Psychol.*, 21, 126-134.

HARTOG, P., RHODES, E. C., and BURT, C. (1936). *The Marks of Examiners.* Macmillan.

HARTOG, P., *et al* (1941). *The Marking of English Essays.* Macmillan.

MCCLELLAND, W. (1942). *Selection for Secondary Education.* University of London Press.

MORRISON, R. L., and VERNON, P. E. (1941). A new method of marking English compositions. *Brit. J. Educ. Psychol.*, 11, 109-119.

RICHARDSON, S. C. (1956). Some evidence relating to the validity of selection for grammar schools. *Brit. J. Educ. Psychol.*, 26, 13-24.

VALENTINE, C. W., and EMMETT, W. G. (1932). *The Reliability of Examinations.* University of London Press.

VERNON, P. E., and MILLICAN, G. D. (1954). A further study of the reliability of English essays. *Brit. J. Stat. Psychol.*, 7, 65-74.

WISEMAN, S. (1949a). The essay in 11 plus selection. *Times Educ. Suppl.*, April.

WISEMAN, S. (1949b). The marking of English composition in grammar school selection. *Brit. J. Educ. Psychol.*, 19, 200-209.

---

* Evidence on this point is given in the forthcoming report, *Selection for Grammar Schools, op. cit.*

(*Manuscript received, September 30th, 1956*).

THE DISTRIBUTION OF PERSONALITY TRAITS IN SEVEN-YEAR-OLD CHILDREN: A COMBINED PSYCHOLOGICAL, PSYCHIATRIC AND SOMATOTYPE STUDY

By MAY A. DAVIDSON, R. G. McINNES AND R. W. PARNELL
(Warneford and Park Hospitals, Oxford.)

SUMMARY.—The physical constitutions of 100 seven-year-old children have been assessed. The more closely correlated facets of physique and behaviour in childhood are reported. A time may come when it will be possible to compare these findings in a series of healthy children with those in a group of children needing psychiatric help. When this is achieved, further light may be thrown on the aetiology of mental illness. In the meantime, it has been shown that the somatotype can give some indication of the kind of behaviour to expect in different individuals, and that exact measurements support what has for centuries been appreciated in less precise form, namely that physical constitution plays an important part in shaping people's lives.

I.—HISTORICAL INTRODUCTION.

EVER since Galen's time men have speculated about the relation between physique and temperament. An early example of such speculation is the Greek theory of humours. This theory held that a person's temperament depends on the the amount that his body contains of certain ' fluids,' such as bile and phlegm. It remained fashionable for centuries and became the stock in trade of poets and dramatists ; but it is only in the present century that a more scientific interest in the precise relation between physique and temperament has been shown. It has long been known, for example, that short fat people are more liable to swings of moods from sadness to elation than others, and that lean people tend to be anxious. But any attempt to develop and utilise the idea of constitutional influences has been retarded by a number of factors. The notion that behaviour is linked with physique shares the unpopularity of determinism ; constitution is difficult to define and to measure ; and the dynamic aspect of human behaviour has recently overshadowed all other considerations.

Yet physique is the material from which a person's style and achievement are fashioned, so its inherent shape and quality are likely to be relevant to the results. If this is so, then the acquisition of a reliable body of data concerning the relation between physique and behaviour would be of importance in many fields, of which psychiatric medicine and child guidance are obvious examples.

Kretschmer (1921) was one of the first to draw our attention to the relation between physique and personality. Sheldon (1942) and his associates have rendered an invaluable service in this field by isolating the main components of physique and drawing up a method of precise observation and measurement ; the application of Sheldon's principles and techniques has already given valuable help to the psychiatrist. This present study is an attempt to modify and extend the use of these methods.

II.—SUBJECTS.

The hundred healthy children who form the subjects of this survey attended the Child Health Survey at the Institute of Social Medicine in Oxford from shortly after their birth. A random group of fifty boys and fifty girls who were within six months of their seventh birthdays were selected as subjects. They were given a series of interviews in which they were somatotyped, and psychologically tested while the mothers were given a psychiatric interview.

III.—METHODS.

(a) *Somatotyping.*

The Somatotype is the best estimate on available evidence of the genetically determined body type which is presumed to remain constant throughout life. It is an attempt to spot the developing patterns of physique and so to identify and classify physical features largely determined by genetic influence. A method of classification will be used here that is based on height, weight, fat, bone and muscle measurements, together with standard three-view photographs.

The final somatotype is given as a three figure index with minimal rating 1 and maximal rating 7 (similar to Sheldon's three figure somatotype). A standard score is in turn allocated to the amount of fat, the amount of bone and muscle development relative to height, and the degree of linearity ; the body with the greatest height and least weight giving the highest index for linearity. These components will be described, following Sheldon, as:

Endomorphy—the component describing roundness of physique and ability to grow fat.

Mesopmorphy—the component describing bone and muscle development.

Ectomorphy—describes the degree of linearity of body build and correlates closely with the index height divided by the cube root of weight.

The somatotype maps below show how this three dimensional continuous distributive classification of physique may be plotted on a two dimensional diagram. They illustrate proximity between persons of similar build, and increasing distance between those of progressively differing body shape. On this triangular map, the fattest person with index 711, or 7 for fat, 1 for bone and muscle and 1 for linearity, is plotted at the bottom left hand corner, the most muscular, with index 171, is plotted at the top, the most linear type, 117, at the bottom right-hand corner. The 444 is plotted at the centre. The majority of findings will be expressed in relation to this basic triangle.

The fifty boys in the sample have been plotted in Fig. 1 and the fifty girls in Fig. 2. Note that the boys' distribution lies largely on the muscular side of the mid-point, or at least on the more muscular side of the line joining type 117 to type 551. Thus, the majority of boys have more muscle than fat in their constitution. On the girls' chart the position is reversed ; the majority of girls have proportionately more fat than muscle.

(b) *Psychological Tests.*

While the mothers were being interviewed by Dr. McInnes, the children were given these tests by Miss Davidson :

(1) The Revised Stanford Binet Test, Form L.
(2) A test of Fluency.
(3) The simplest version of the Progressive Matrices (Set A, AB, B 1947).
(4) Schonell's Graded Word Reading Test.

At a subsequent interview, the childern were seen by a second psychologist, Mrs. Field, in their homes, where they were given the Wechsler Intelligence Scale for Children (WISC). There were two reasons for giving this test ; firstly to contribute to a standardisation for English children aged 7 years ; and secondly, as it comprised both verbal and performance tests, to examine the hypothesis that children scoring highly on the verbal part of the test might differ constitutionally from those scoring highly on the performance tests. This hypothesis has received some support from the findings, as will be seen later. At a further interview, again in home surroundings, Mrs. Field gave the children a battery of projective tests which included the Rorschach Ink Blots (also with a view to standardisation of age 7), two cards from Bellak''s Thematic

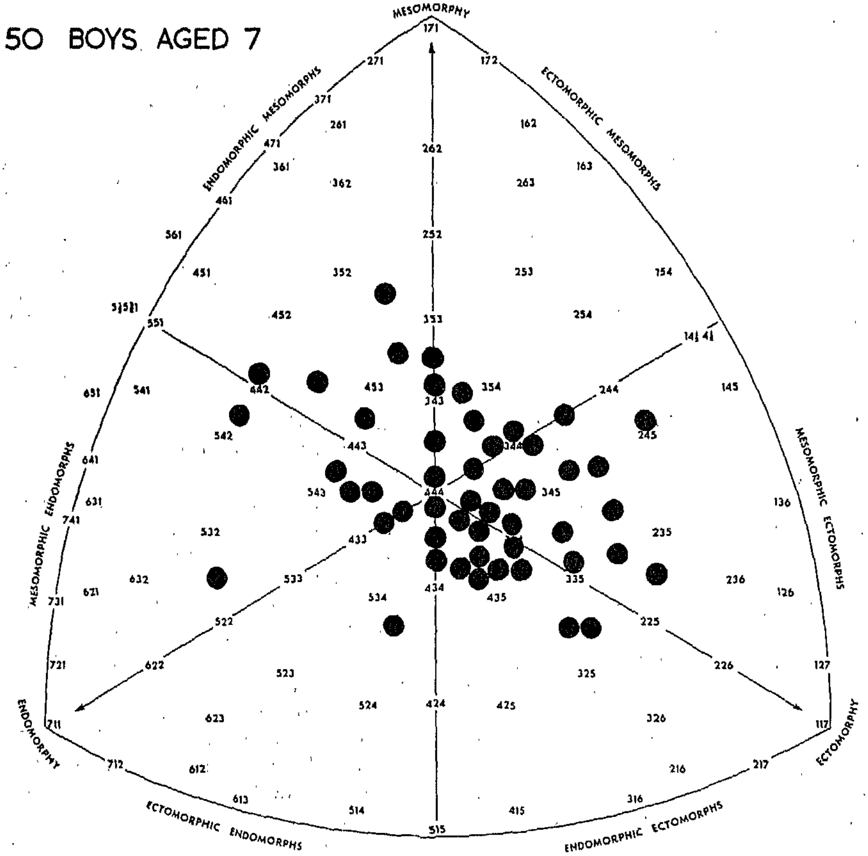


FIG. 1

Apperception Test for Children (CAT), and selective silhouettes from the test devised by Mrs. Drysdale of the Mulberry Bush School and Mr. Babington Smith of the Oxford University Institute of Experimental Psychology. Ninety-five out of 100 children completed all these tests.

(*c*) *Psychiatric Interview.*

A form of inquiry was devised as the basis for interview with the child's mother. It covered the main events, activities and influences of the first seven years of the child's life. The first page dealt with birth, feeding history and toilet training, the second with the child's attitude to parents or parent substitutes—to receiving, to possessions, to giving and to frustration or deprivation. Each of these headings carried its own appropriate list of five possible answers, : e.g., under the heading "attitude to frustration" were listed acceptance, anxiety, depression, anger and withdrawal. The third page of the inquiry dealt in greater detail with the attitude of the child to siblings, other children, teachers, and adults in general, and with the child's reaction to loss of its mother or separation from her. It dealt also with the child's illnesses and

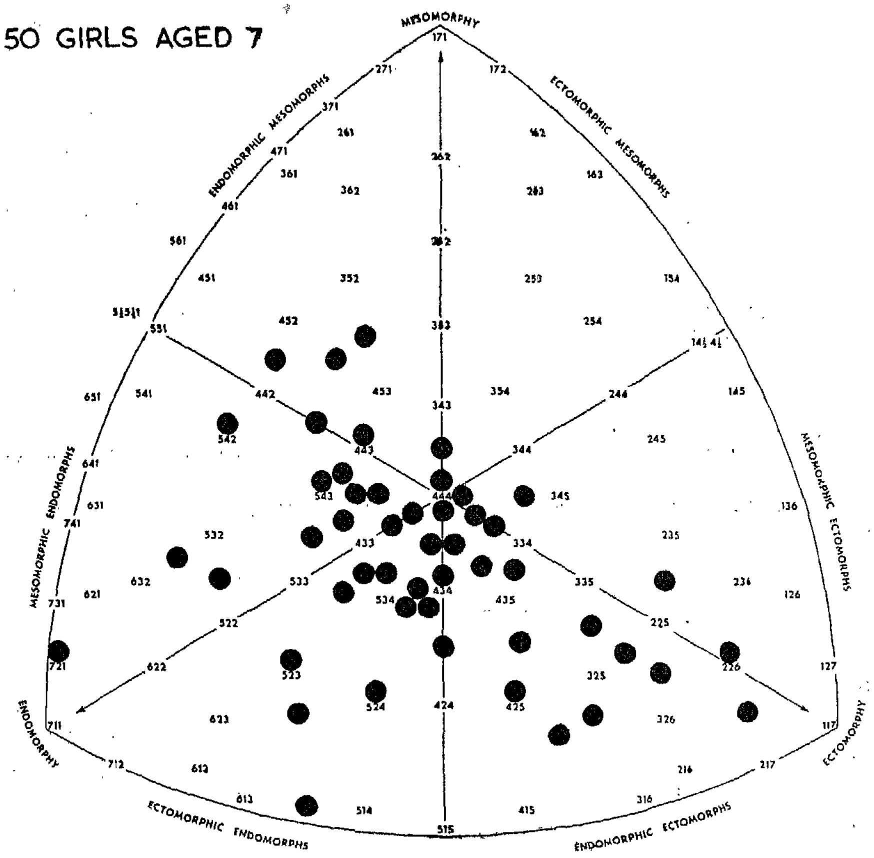

50 GIRLS AGED 7

FIG. 2

his reactions to them, and with the usual milestones in development. There followed an inquiry into the attitudes of mother and father to their own parents, to each other, to the child, and to their other children, if any.

The last section of the survey recorded (on a three-point scale) the presence or absence of thirty-three personality traits at age seven. It was noted whether the child was anxious, conscientious, suspicious, fussy, aggressive, showing-off, deceitful, shy, dreamy, self-absorbed, and so on.

Finally, there was added a check list of thirty-six of the common signs and symptoms of emotional unrest in children, such as bed-wetting, wandering, nail-biting, nightmares, sleep-walking, thumb-sucking and others, all rated on a three-point scale.

It should be realised that this survey was not simply a questionnaire, but the basis of an interview with the child's mother, generally lasting for about 45 mintues to an hour, and conducted in every case by Dr. McInnes, who also saw the child briefly, mainly to satisfy himself that there was no obvious discrepancy between the information given and the observations made.

True, the recordings and ratings entered in the inquiry were in many cases not a record of demonstrable facts, but rather represented the opinion of the interviewer on the information given; nevertheless, the inquiry was standardised, as far as seemed consistent with an interviewing technique, and it had the advantage of a single observer throughout.

(*d*) *Correlation Techniques.*

In the material to be described tetrachoric correlations have been used and co-efficients exceeding 0·27 have been selected for mention; this gives a convenient level of significance (P approx. ·10) for a sample of fifty, and all traits have been correlated with midway cut-off points in the physical component distributions. Most of the correlations fall between 0·30 and 0·70. The diagnostic and prognostic value of such coefficients for individual cases is small, but a fairly large number of moderate values is a guide to their usefulness in close conjunction with the result of psychological testing and psychiatric interview.

IV.—RESULTS.

(*a*) *Cognitive Test Findings.*

Figure 3 shows the distribution of test scores for the hundred children on the Revised Stanford Binet Scale. The group is skewed towards the bright side. The mean I.Q. is 108, and there is only one child with an intelligence quotient below 85.

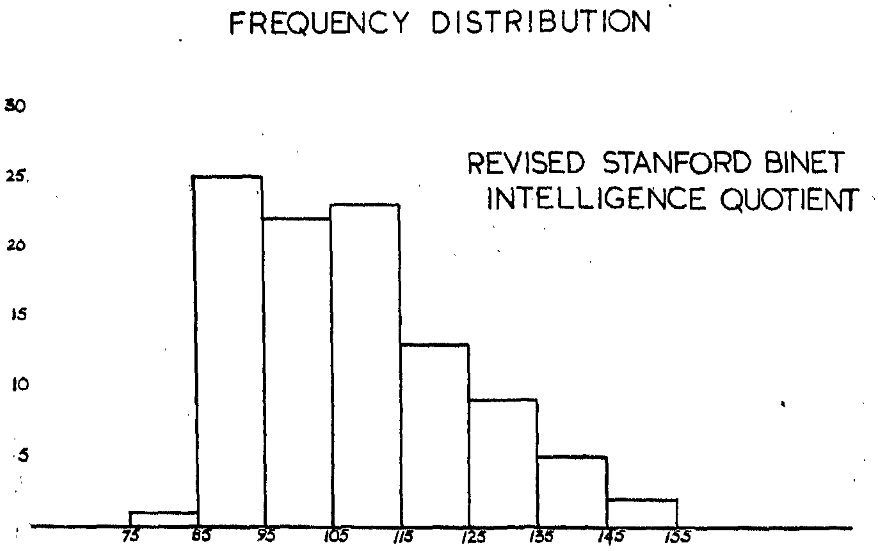
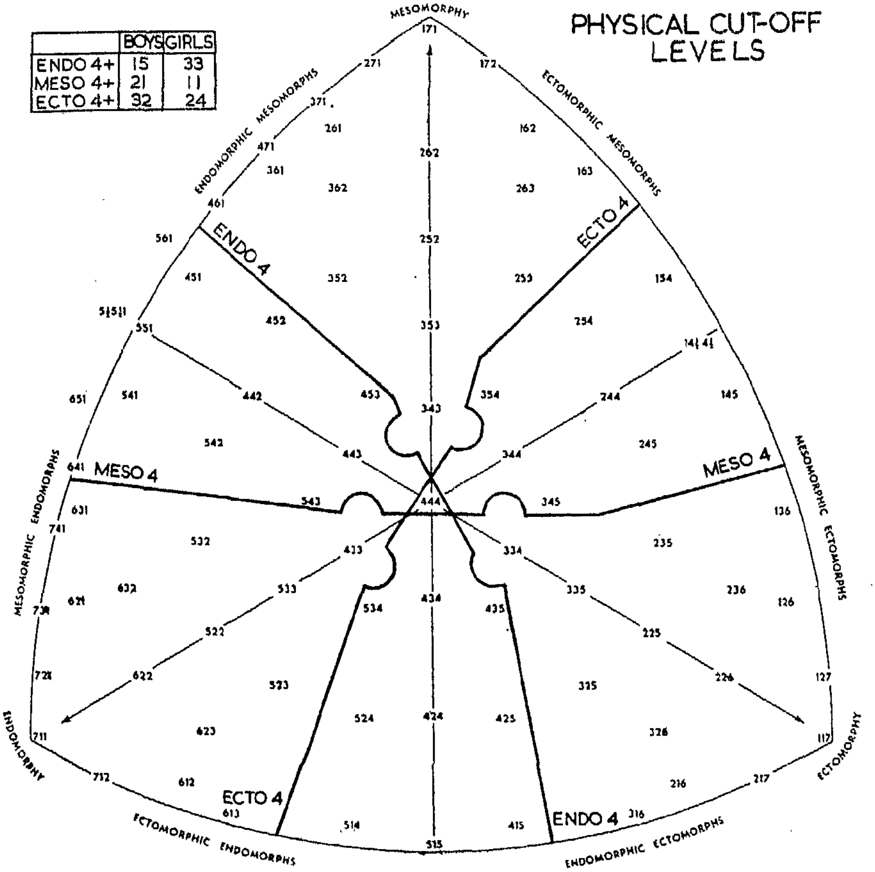


FIG. 3

Most of what follows is confined to the group of fifty boys. The results for the girls are less informative because they are a much more widely distributed group with regard to intelligence. Many of the girls came from private schools, while the boys came mainly from elementary school and are, therefore, a more homogeneous sample.

For present purposes the endomorphic group are those with a score of 4+ in endomorphy, the mesomorphic group have a score of 4+ in mesomorphy, while the ectomorphic group have a score of 4+ in ectomorphy (Fig. 4). Table 1 gives the correlation co-efficients between physical components and above-average Binet test scores.



	BOYS	GIRLS
ENDO 4+	15	33
MESO 4+	21	11
ECTO 4+	32	24

FIG. 4

TABLE 1

	Boys	Girls
Endomorphy 4+ Fat	0·31	0·32
Mesomorphy 4+ Muscular	0·01	0·00
Ectomorphy 4+ Linear	0·19	0·22

It may be seen that endomorphs of either sex obtain the highest scores on this type of intelligence test. Ectomorphs come next and mesomorphs do least well. The superiority of endomorphs is statistically significant, though at a low level.

There are two further points which deserve mention. The first is that many of these children (54 boys, 43 girls) had been previously tested on the Binet Scale at the age of five years. A comparison of the two sets of results shows that there is a tendency for mesomorphs of either sex to obtain lower scores on the second test (r= -0·29) while the endomorphic boys tend to maintain their level or even improve (r=+0·30). The second point is that a hundred male undergraduates who had volunteered to act as controls in a study of mental breakdown in Oxford University students, provide evidence of a similar relation between somatotype and level of achievement on a group intelligence test (correlations are endomorphs +0·25, mesomorphs 0·00, ectomorphs +0·03).

The Wechsler Intelligence Scale for Children (WISC) consists of a verbal and a non-verbal battery ; the two contribute equally to the overall result.

The somatotype of the boys was correlated with above and below average scores on the non-verbal battery. Here mesomorphs did best (+0·15), and endomorphs least well (-0·03). Next, the somatotype distribution of those obtaining higher verbal than non-verbal scores was examined. This time mesomorphs did badly (-0·35), ectomorphs well (+0·34), while endomorphs were average (-0·06).

The scholastic attainment of children of both sexes was compared with their body types ; scholastic attainment being assessed on the basis of reading level. The number of words read on the Schonell Graded Reading Test was compared with the somatotype, and it was found that ectomorphs of either sex were by far the best readers (+0·35 and +0·47 for boys and girls, respectively) ; endomorphs were in the middle (+0·10 and -0·10) while mesomorphs did least well (-0·10 and -0·47).

Eliminating the influence of intelligence on reading level (by comparing reading age with mental age on the Revised Stanford Binet), it was found, taking children with reading quotients of above and below ninety, that once again it was the ectomorphs who did best (boys +0·22, girls +0·25) ; the endomorphs were in the middle position (-0·08 and +0·18), while the mesomorphs did badly (-0·41 and -0·43). These results suggest that the ectomorphs were putting more drive into acquiring their reading skill than the endomorphs or mesomorphs.

It is worth mentioning that in the Annual School Examination at 11+ the general picture is much the same, according to a survey carried out by Dr. Westropp. For example, mesomorphs do less well in English (-0·34), a similar result being obtained even when due allowance is made for the influence of intelligence on this type of examination. Moreover, Parnell, in his survey of the undergraduate population sponsored by the Institute of Social Medicine, found that a large proportion of those taking literary subjects like English and History were ectomorphic.

The conclusions to be drawn from the cognitive test findings are only tentative, as the numbers in the groups are small and the correlations would certainly need to be checked on a much larger group. However, the provisional findings suggest that there is some association between ectomorphy and verbal ability on the one hand and between mesomorphy and practical ability on the other. Yet it is the endomorphs—" all rounders " in more than one sense—who do best on a test of the Binet type, which involves arithmetic as well as a mixture of verbal and non-verbal items. These observations, if correct, have an important bearing on the problem of educational selection. It would appear that the 11+ examination, which essentially calls for verbal rather than practical ability, must favour ectomorphs against mesomorphs, with the endomorphs occupying a neutral position.

To illustrate the point the results from 405 boys and 461 girls taking the 11+ examination in 1954 are shown in Table 2 ; the following correlations were found with obtaining a grammar school place.

TABLE 2

	Boys	Girls
Endomorphy 4+ Fat	+0·10	−0·01
Mesomorphy 4+ Muscular	−0·25	−0·13
Ectomorphy 4+ Leinar	+0·10	+0·22

In view of this, it seems odd that mesomorphs who reach the university should (according to Parnell's Survey) be relatively successful in their final examinations, obtaining the expected number of first-class honours, the highest porportion of seconds, and the lowest of thirds. It may well be that those mesomorphs who reach the university are exceptionally gifted intellectually, or else possess some helpful characteristic of temperament, such as drive, which assists them in the liberal atmosphere of Oxford where they seem to do so well.

(*b*) *Projective Tests.*

In the case of the Rorschach test the child was simply asked to say what the ten ink blots looked like to him. In the case of Bellak's TAT cards the child had to respond by telling a story. In interpreting the findings from these less objective tests, the examiner's clinical impression of the child was taken into consideration.

To deal first with the quantitative aspect of the Rorschach findings ; if sheer productivity is considered, i.e., the total number of responses given, mesomorphs are by far the most productive (endos +0·15, mesos +0·38, ectos −0·05).

On the basis of accepted Rorschach theory, responses may be classified into three main types according to whether they are determined primarily by form, by colour or by the impression of movement. Attention to the form of the blot is held to reflect intellectual control, attention to colour reflects emotional response to the outside world, while a capacity for seeing movement in the blots is regarded as a measure of the subject's inner life. The relation between type of response and somatotype is given here :

Movement ..	Endos r +0·05	Mesos r +0·10	Ectos r +0·22
Form ..	Endos r +0·11	Mesos r −0·20	Ectos r +0·02
Colour ..	Endos r +0·23	Mesos r −0·06	Ectos r −0·12

None of these results is statistically significant, but they do at least show certain general tendencies of the kind which Sheldon's description of temperament in adults might lead one to expect. To give some examples, ectomorphs have most intellectual control and mesomorphs least. This agrees with findings on the cognitive tests. Emotional response to the environment goes in the expected order of endomorphy, mesomorphy, ectomorphy, and the inner imaginative life is richest in the ectomorphs.

The remaining results are purely qualitative. The effective use of intelligence on these projection tests may be judged from various indications. The evidence here points to the endomorph as better able to use his intelligence, when confronted with this cryptic situation, than his mesomorphic fellow.

The unfortunate ectomorph is forced back into his protective shell and is, therefore, unable to function at his best.

The results of projection tests are notoriously difficult to interpret, but the present findings suggest a demonstrable relation between test results and somatotypes. To the ectomorph his inner world is more important. He appears, therefore, as less able to respond to his environment, less adaptable socially and less able to use his intelligence when confronted with a new and non-academic situation, despite his frequently superior performance in the academic sphere. To the endomorph the external environment seems to provide the stimulus, with the result that he is affectively lively, socially well adjusted and better able to cope with new situations. The mesomorphs are in the middle of the road. The results agree reasonably well with the cognitive tests, which showed the ectomorph putting drive into academic achievement, perhaps because he has fewer resources for gaining satisfaction in the outside world.

It is necessary to emphasize once again the extremely tentative character of the present findings. Although further investigation is clearly called for, the available evidence suggests that the kind of association between personality

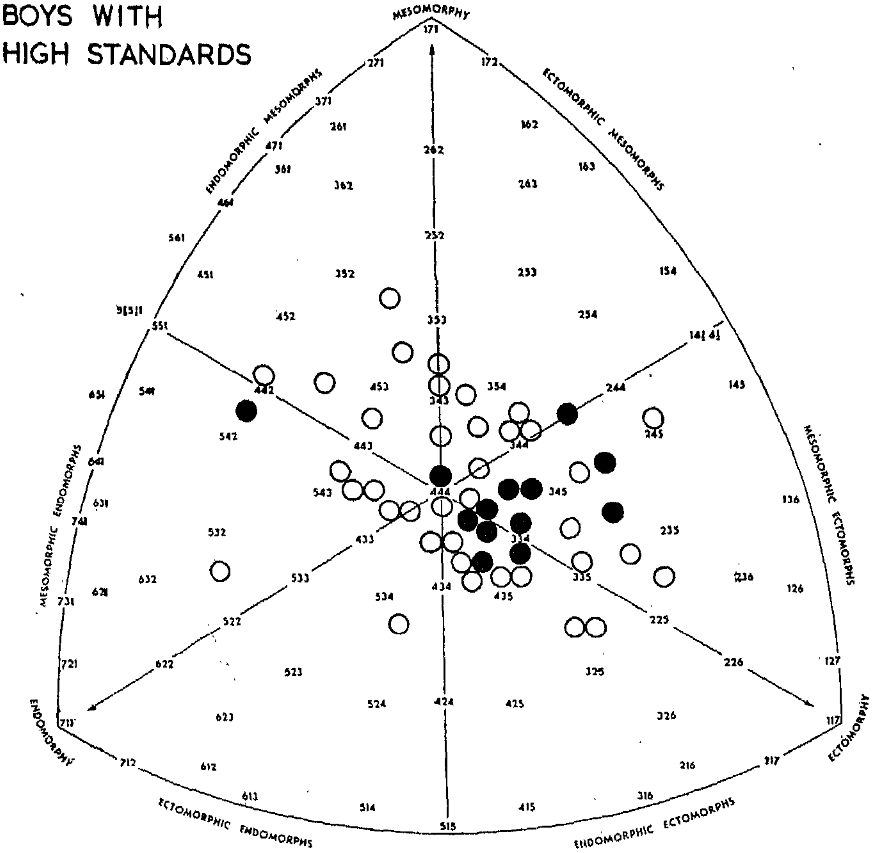


FIG. 5.

and physique which Sheldon observed in an adult population is already virtually established at the tender age of seven.

(c) *Psychiatric Assessments.*

The following examples illustrate the relation between Physique and various Traits of Behaviour.

(i) *High standards and meticulousness.*—The somatotype distribution of boys described by their mothers as having habitually high standards is shown in Fig. 5. Fig. 6 shows the distribution of girls.

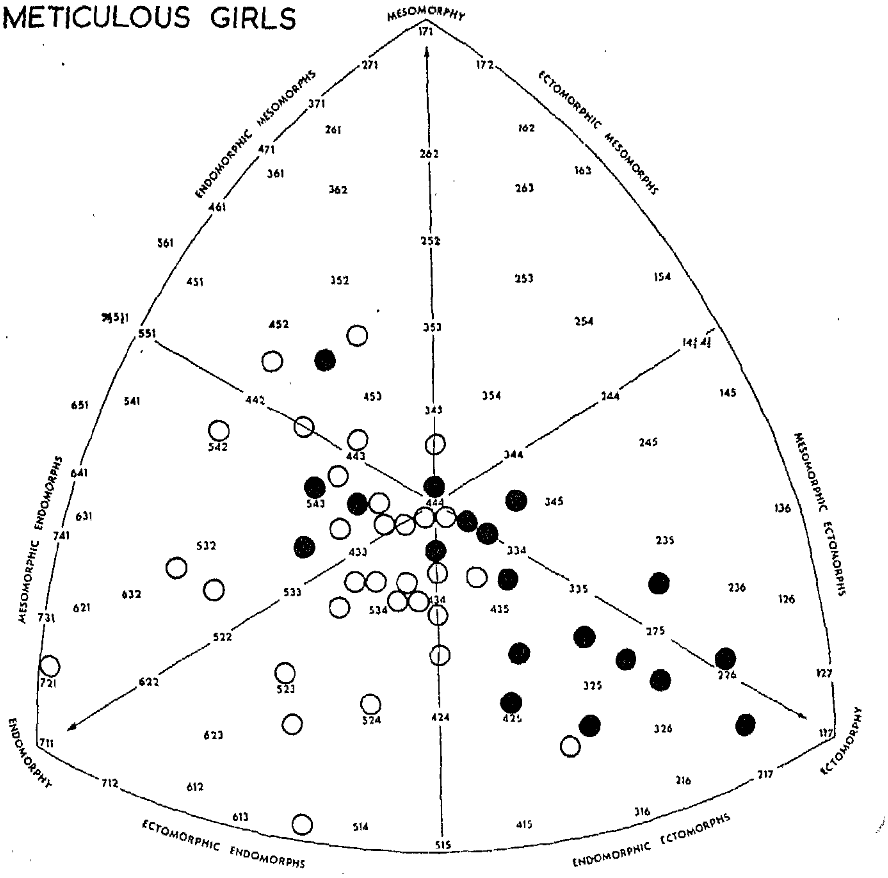

METICULOUS GIRLS

FIG 6.

The boys will be seen to congregate just east of the centre. Their build is a little lighter than average; their muscularity is not diminished and the lighter build is achieved at the expense of fat. With one exception, there is a notable absence of boys with heavier than average build.

Turning to the distribution of girls with high standards, 19 out of 50 were so described and most were linear in build. Only 4 were not linear, though it would be reasonable to expect to find 10 such. It seems, therefore, that boys and girls with high standards tend to be more slender in build than average. As a counterpart to this finding in children, it may be mentioned that delinquent young men and women of an age suitable for Borstal training have been previously reported in quite separate surveys as more often stocky or tubby in build (Glueck and Glueck, 1950 ; Sheldon, *et al,* 1951 ; Epps and Parnell, 1952).

(ii) *Anxiety.*—In this survey there were 23 anxious boys, of whom 16 were ectomorphs, giving a prevalence of 69·6 per cent., compared with 25·9 per cent. among 27 non-ectomorphic boys. It is also noteworthy that only one anxious boy was a mesomorph, leaving 9 mesomorphs in the general distribution who were not anxious.

(iii) *Emotional Unrest.*—Table 3 gives the 10 most common signs of emotional unrest.

TABLE 3

	Boys n=50	Girls n=50	Total Prevalence
	No.	No.	%
Thumb-sucking	18	24	42
Nail biting	17	21	38
Temper tantrums	22	15	37
Teeth grinding	18	16	34
Nightmares	16	15	31
Panics	9	11	20
Breath holding	8	10	18
Bed wetting	8	8	16
Night terrors	9	6	15
Head banging and/or rolling	7	4	11

Among 14 boys with a high degree of disturbance (i.e., with 4 or more of the above signs), ectomorphy was dominant in all but two. Indeed, the correlation of central ectomorphy with high scores for emotional unrest gives the significant coefficient of +0·65.

In addition, among these healthy children, 42 per cent. had sucked their thumbs, noticeabaly at some time in the first seven years, 38 per cent, bit their nails, 37 per cent. had temper trantrums, 34 per cent. ground their teeth at night, 31 per cent. had nightmares. That is to say, between one-third and one-half of healthy children do these things. This is perhaps surprising, but it may also be reassuring, especially to mothers who are over-sensitive to the occurrence of such behaviour.

This table suggests that it may be necessary to revise our ideas of the normal and the abnormal in the behaviour of children. Items of behaviour which occur in nearly one-half of a group of healthy children can hardly be called pathological, unless they occur with exceptional intensity or persistence.

(iv) *Submissiveness.*—The chart in Fig. 7 shows the distribution of those boys and girls who were notably submissive. They are almost all in the lower half of the chart—in other words it seems that submissiveness tends to be associated with a low rating in mesomorphy. The submissive group comprises 24 children, 11 boy and 13 girls, yet none of the 13 mesomorphs appeared as submissive. Seven of 11 submissive boys were ectomorphs.

It is an interesting fact that 3 of the most ectomorphic girls in the general distribution appear in this submissive group (in the bottom right-hand corner) ; and that the 2 most markedly endomorphic boys are also in this area of the chart (mid-bottom left).

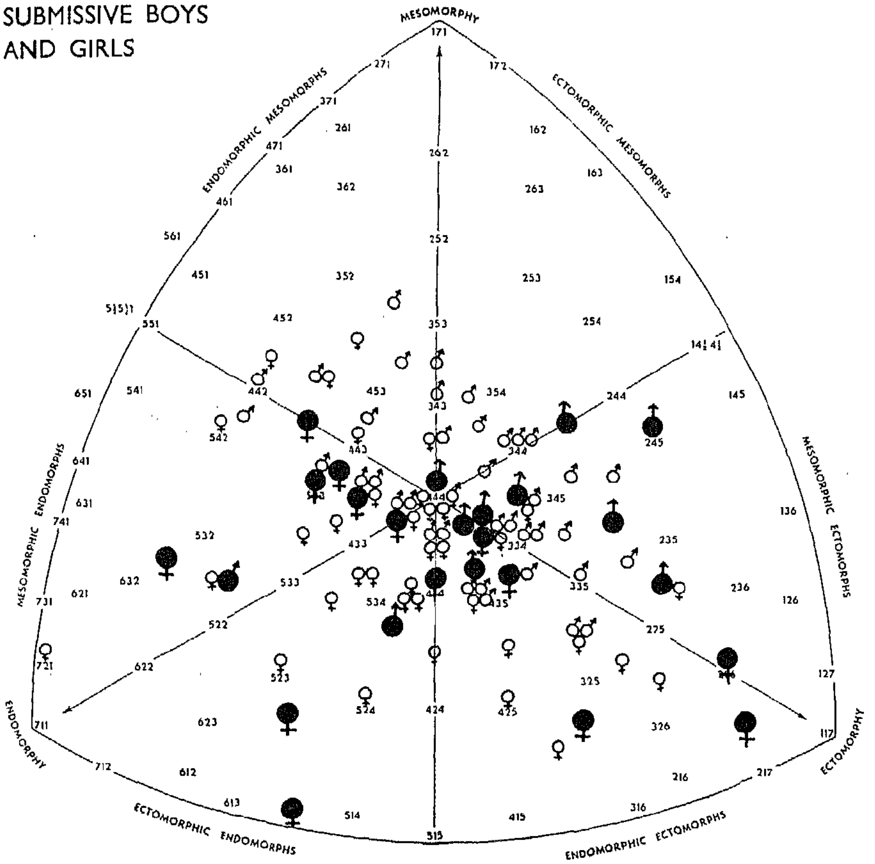


FIG. 7.

This would seem to illustrate the well-known observation that people submit either because they are easy-going by nature, or because they are too speculative to reach a decision on their own ; in either case there is a relative absence of thrust in their constitutional make-up. Though this might be regarded as obvious, the point to be made is that, in the management of children, a personality trait such as submissiveness may be misinterpreted as to its origins, incorrectly estimated as to its extent, and, therefore, perhaps wrongly handled by parents, teachers and doctors, unless the constitutional factor be taken into account. People responsible for children would all like to be able to distinguish more certainly between the child who submits because he just does not care, and the one who submits because he cares too much. It is not suggested that somatotyping can yet produce infallible answers to such questions, but even at 7 years, it seems that it may be able to give us a firmer indication of what kind of child we are dealing with, of what not to expect from him, and of what would seem to merit careful cultivation.

(v) *Difficulty in Communicating.*—This trait is not quite the same as shyness or self-absorption, although it is often associated with them. The child with this trait may overcome initial shyness and come out of his shell, only to find himself still puzzled and at sea with his associates because he cannot discover a means of " putting himself across " to them, or of grasping fully their personal meaning for him.

In this survey 9 boys and 7 girls possessed this characteristic ; none appear in the mesomorphic zone. It seems, therefore, that for ease of communication it is best to be on the western side of the midline.

(vi) *Nightmares and Night Terrors.*—Nineteen boys had nightmares and night terrors and 12 of these were predominantly ectomorphic ; in fact, 52·1 per cent. of ectomorphs had nightmares, compared with 25·9 per cent. of those with other builds. Although this group is small, it is perhaps worth noting that none appears in the endomorphic zone. Only 3 mesomorphs had nightmares, i.e., one-third of the total of 9 in the general distribution.

(vii) *Suspiciousness and Resentfulness.*—Eight boys showed this trait and again they appear mainly in the ectomorphic area of the chart. There is a noteworthy absence of endomorphs in this small group.

In boys with less than average muscle, anxiety and " showing off " were amongst the highest correlates. Confidence and well directed restlessness were more usual in fatter boys, ability to communicate and aggressiveness more common in the stocky ones.

Among the more significant negative correlations, more muscular boys were noted as showing little anxiety ; in addition they were not fussy or shy and seldom tended to dramatise themselves and their affairs. Fat boys were neither suspicious nor withdrawn and had no difficulty in communicating their wishes and feelings to others. They were not submissive but neither were they resentful or dreamy. The boys with less than average muscularity, though not necessarily meek, seldom had explosive rages or similar behaviour. Boys of slender build showed, according to their mothers, rather less confidence ; they seemed too, if anything, less affectionate.

Among girls the strongest positive correlations were found with ectomorphy; slender girls were notably meticulous, fussy and conscientious, though also apt to appear on the defensive. Negative correlations, a longer list at a somewhat lower level, showed in the more muscular girls absence of shyness, dreaminess, and odd or fantastic behaviour. Thickset girls were less anxious, less inclined to be suspicious, while tubby girls were seldom sad, seldom on the defensive, were disinclined to blame others and had little difficulty in communicating or getting on with other people. Lastly, among those with less than average muscle incorrigibility was seldom manifest.

V.—DISCUSSION OF FINDINGS.

These findings suggest that, even at the age of 7, there is a discernible association between behaviour pattern and measurable somatic components. Because these *are* measurable, and because the correlations with many traits are significant, this statement goes substantially beyond the general assertion that fat children tend to be easy going, and thin children tend to be anxious.

The findings seem to offer reasonable hope that the somatotyping of children at the age of 7 will make it possible to pick out those who are especially vulnerable. It may also prevent people from expecting the impossible—it is unlikely, for example, that a child with a low rating in mesomorphy can be made into an athlete, and it is probably unwise to try. The anxious mother of an endomorphic child may possibly derive some consolation from realising that her failure to make the child as careful as she would like arises, not from her own shortcomings, but from the fact that the child is just not built that way. Parents may feel more justified than they do in treating their different children differently, and teachers may take fresh comfort from this confirmation of their suspicion that they cannot impress all of the pupils all the time.

From the psychiatric point of view the most important hope arising from this work is that of an increased ability to prevent breakdown and maladjustment. Too often psychiatrists are trying to patch up after the damage has been done. Even in child guidance work, the helpful measures often have to be applied after the event. Moreover, the ability to discern proclivities from the measurement of somatic components would discourage too exclusive a consideration of historical, social and environmental influences as causes of breakdown.

The work already undertaken ,or in porgress, in this field suggests possible lines of research for the fufture. Studies being made at the Warneford Hospital on the physique of men with a disposition to mental illness show that patients in need of psychiatric care, who have particular symptoms, tend to have physiques of a characteristic kind ; further that the natural course of mental illness, its severity, and the patient's response to such treatment as electroconvulsion therapy and insulin, are intimately tied to physique ; and, lastly, that electroencephalographic findings may also relate to body build.

Somatotyping could be used to assist in the study of interpersonal relations. It would provide an almost unlimited variety of experiments. Perhaps it would be best to begin with family studies. These might not only throw fresh light on the reasons for family tensions, but make a useful contribution to the study of genetics.

Many other applications of somatotyping in relation to behaviour can be thought of. Proneness to physical diseases, occupational aptitude, selection procedures, response to climatic changes, are only a few of the areas of enquiry which might be found fruitful.

* * * * * *

ACKNOWLEDGEMENTS.—The 100 healthy children in this survey attended the Child Health Survey of the Institute of Social Medicine in Oxford from shortly after their birth. The authors, therefore, wish to thank not only the parents for their co-operation over the years, but also the Director and members of the Institute staff. Special thanks are due to Dr. C. K. Westropp who took many of the physical measurements and to Mrs. M. Blaschko who arranged for the children to attend for photography and psychological tests, and for the mothers to be interviewed. Mr. B. Babinton Smith kindly provided some expert statistical advice ; however, responsibility for the statistical treatment employed rests with the authors.

VI.—REFERENCES.

BELLAK, L. (1949). *Children's Apperception Test* (C.A.T.). New York : C.P.S. Co.

EPPS, P., and PARNELL, R. W. (1952). Physique and temperament of women delinquents compared with women [illegible] *J. Med. Psychol.*, 25, 249-255.

GLUECK, S., and GLUECK, E. [illegible] *Juvenile Delinquency.* New York : Commonwealth Fund.

KRETSCHMER, E. (1921). *Körperbau und Charakter.* Berlin : Springer.

PARNELL, R. W. (1953). Physique and choice of faculty. *Brit. Med. J.*, ii, 472-475.

PARNELL, R. W., PEIERLS, G. E., STANDLEY, C. C., and WESTROPP, C. K. (1955). Physique and the grammar school entrance examination. *Brit. Med. J.*, i, 1506-1510.

SCHONELL, F. E., and SCHONELL, F. J. (1950). Graded word reading test. In *Diagnostic and Attainment Testing.* Edinburgh : Oliver and Boyd.

SHELDON, W. H., HARTL, E. M., and McDERMOTT, S. (1949). *Varieties of Delinquent Youth.* New York : Harper.

SHELDON, W. H., and STEVENS, S. S. (1942). *The Varieties of Temperament.* New York : Harper.

SHELDON, W. H., and TUCKER, W. B. (1940). *Varieties of Human Physique.* New York : Harper.

TEMPERAMENTAL CHARACTERISTICS RELATED TO DISPARITY OF ATTAINMENT IN READING AND ARITHMETIC

By R. LYNN
(University of Exeter).

SUMMARY. 1.—Several studies suggest that a lack of balance or disparity in attainment in reading and arithmetic in school children is related to certain temperamental traits. These studies suggested the hypothesis that there is a general relationship between *anxiety* and good reading in relation to arithmetic.

2.—This hypothesis was investigated by studies of the relation of anxiety to reading-arithmetic disparity among a group of primary school children and among a group of secondary modern school boys. In both groups there was a tendency for anxious children to be better at reading than at arithmetic.

3.—This tendency might be due to the greater amount of time which anxious children spend on reading as a way of dealing with their anxieties and finding satisfaction in fantasy.

I.—INTRODUCTION.

THE child who is good at reading but poor at arithmetic is generally recognised as odd. Disparity between reading and arithmetic attainment has frequently attracted the attention of psychologists, and there is evidence from several sources suggesting that such disparity is related both to the total personality of the child and to his home environment. This paper reports the results of an investigation into the relation of levels of attainment in reading and arithmetic to temperamental characteristics and in particular to feelings of anxiety.

II.—REVIEW OF THE LITERATURE.

There appear to be no studies of what might be termed 'attainment disparity' among normal school children. There are, however, several types of study which have an oblique relevance to the question.

(1) Attainment Disparity in Clinical Groups. There is some evidence that neurotic children tend to be better at reading than at arithmetic and that psychopaths are better at arithmetic than reading. In Burt's (1937) view, children who are anxious tend to be better at reading than at arithmetic. Speaking of anxious children, he writes :

> " With most, however, it is arithmetic that offers the severest trials. They would readily echo the sentiments of Pet Marjorie, the seven-year-old friend of Sir Walter Scott : 'I am now going to tell you the trouble and wretched plague that my multiplication gives me. You can't conceive it. The most Devilish thing is 8 times 8, and 7 times 7 is what Nature itself can't endure '." (p. 551.)

Jastak (1941) writes that " high reading and low arithmetic scores tend to occur in abnormal mental states of developmental nature and of long standing as in neurosis and schizophrenia." Feinberg and Moscovitch (1956) report that in most emotionally disturbed children achievement in arithmetic is significantly depressed.

In a study of maternally overprotected children, Levy (1943) found that these children were good at reading but poor at arithmetic. Since it is often thought that maternal overprotection is associated with nervousness in children, this evidence might be regarded as supporting Burt's thesis. However, the question of how parental attitudes are related to attainment disparity is a separate problem and will not be pursued here.

Both Chase (1932) and Feinberg (1947) report that juvenile delinquents are better at reading than arithmetic.

Very little research has been published concerning the relation of temperamental characteristics of good arithmetic in relation to reading. According to Jastak (1941) this pattern is frequently found in patients with acquired psycholosis due to alcoholism and syphilitic infection.

(2) Attainment Disparity and Sex Differences. There is a considerable literature showing that girls tend to be better than boys at reading, but poorer at arithmetic (e.g., Ayres, 1909; Lincoln, 1927; Stroud and Lindquist, 1942; and many other papers).

(3) Temperament Differences in Intelligence Test Scatter. Closely related to the question of reading-arithmetic disparity is that of the scatter of verbal and perceptual or performance sub tests in intelligence testing. Several studies of the relation of temperament to test scatter are available and these tend to be consistent with the findings on reading-arithmetic disparity. Himmelweit (1946) found that obsessional and anxiety neurotics have a high verbal ability in relation to their performance intelligence, and these results have been confirmed by Eysenck (1947, p. 124). Similar findings are reported by Rapaport (1946) and Altus and Clark (1949). On the other hand, especial weakness in verbal intelligence has been noted among delinquents and psychopathic personalities (Wechsler, 1941; Piquer y Jover, 1946; Bender, 1950).

Taken as a whole, these studies suggested the hypothesis that good reading in relation to arithmetic might be associated with nervousness and in particular with anxiety. The following investigation was undertaken to test this hypothesis.

III.—THE INVESTIGATION.

(*a*) *Primary School Children.*

Subjects.—The subjects were 80 unselected normal children attending a primary school. There were 42 boys and 38 girls, with an age range of 7·5 to 11·0 years and a mean age of 9·6. The mean I.Q., as assessed by the Stanford-Binet test, was 112·2, with a standard deviation of 17·2. The school was situated in a predominantly middle class district, and this probably accounts for the rather high mean intelligence.

Attainment Tests.—Reading attainment was assessed by Schonell's Graded Reading test, using the norms as modified by Vernon (1950a). The mean reading age was 9·11 (S.D. 20 months). Arithmetic attainment was assessed by a mechanical arithmetic test devised by the local education authority. The mean arithmetic age was 9·8 (S.D. 17 months). A ' disparity score ' was derived by subtracting the arithmetic age from the reading age.

Anxiety Tests. The children were given a number of tests of anxiety and the scores of these tests were positively inter-correlated. Two of the tests are described here and their relation to attainment disparity shown below.

(i) *Personal Anxieties.*—This is a test comprising six questions about the child's anxiety in his personal relationships, especially those concerning his mother. For example, the children were asked whether they had ever been afraid that their mother might die. The six questions are shown in full elsewhere (Lynn, 1955). Table I shows the relation of anxiety to attainment disparity in a contingency table. Chi squared=15·3, and is significant at the 5 per cent. level.

(ii) *Himmelweit and Petrie's Test.* This is a projection test devised by Himmelweit and Petrie (1951) for the study of children with behaviour disorders. The test consists of five brief stories in which the child reads of a social dilemma which must be dealt with in one of five ways. The child reads the story or hears it read to him, and then puts a tick against the response which

TABLE 1

READING-ARITHMETIC DISPARITY.

−2 −1 0 +1 +2

Anxiety	0	12	5	11	5
	1–2	4	10	5	8
	3–5	3	2	5	10

he thinks 'right.' Responses were scored anxious if they suggested that the child was seriously perturbed by the rejection or disapproval of other children or of the teacher.

Table II presents the results of this test and their relation to disparity in a contingency table. Chi squared=14·7, and is significant at the 5 per cent. level.

TABLE 2

READING-ARITHMETIC DISPARITY

−2 −1 0 +1 +2

Anxiety	0	10	6	10	6
	1	7	9	6	5
	2–4	2	2	5	12

(*b*) *Secondary Modern School Boys.*

The Subjects.—The subjects were 45 normal boys attending a secondary modern school. They were all the boys due to leave the school at the end of the term in which the testing took place. The age range was from 14·6 to 15·6.

The Attainment Tests.—Reading attainment was assessed by Schonell's Graded Reading Test using Vernon's modified norms ; the mean Reading Age was 13·6 (S.D. 14 months). Arithmetic attainment was assessed by Cattell's Mechanical arithmetic test with the addition of some standardised harder problems to make the test suitable for 15-year-old boys ; the mean Arithmetical Age was 13·7 (S.D. 22 months). A 'disparity score' was derived by subtracting the Arithmetic age from the Reading age.

The Anxiety Test.—Anxiety was assessed by a modified version of the Taylor Anxiety Scale (Taylor, 1953). Of the 50 questions, 20 were asked and the rest omitted as unsuitable for English school children. Using Taylor's notation, the questions asked were Nos. 1, 4, 7, 11, 12, 14, 15, 16, 18, 19, 25, 27, 28, 32, 35, 38, 41, 44, 47, 50. These are questions about anxiety symptoms such as worrying, poor sleep, nightmares, feeling embarrassed easily, feeling depressed and unhappy and the like.

Procedure.—The boys were interviewed individually in the school for ten to fifteen minutes each. After some informal conversation about school, their prospective jobs, etc., the Taylor questions were asked and, if necessary, explained to ensure that they were undertsood.

Results.—Table III shows the relation of Taylor Anxiety scores to disparity in a contingency table. Chi-squared=2·6, which is not significant at the 5 per cent. level, but the tendency for anxiety to be associated with good reading—poor arithmetic is again present.

TABLE 3

READING-ARITHMETIC DISPARITY

		—3	—2	—1	0	+1
Anxiety	0–3	6	2	1	5	
	4–6	7	5	2	6	
	7–10	2	2	2	5	

DISCUSSION OF RESULTS.

(1) The positive association of anxiety with better reading than arithmetic is consistent with the clinical studies of attainment disparity and suggests that these should be viewed as extreme cases of a general relationship which is found among normal school children. The results are also consistent with the high verbal ability found in anxiety neurotics, and suggest that there may be a general association between anxiety and high VEd : low KM factors.

(2) The association of anxiety with poor reading in relation to arithmetic may throw some light on the problem of specific dyslexia, the specific backwardness in reading which is supposed by some writers to be due to a neurological defect or to differential maturation in the brain. It is, however, widely recognised that this condition is frequently accompanied by emotional and temperamental maladjustments (e.g., Schonell, 1942), and where this is so the present results suggest that insufficient anxiety may be a relevant temperamental trait. It is possible that some children do not learn to read because they are not anxious enough. If this is so, evidence for the inheritance of specific dyslexia (Hallgren, 1950) could be attributed to the inheritance of low anxiety, not poor reading ability as such.

(3) Both the present and previous findings could be explained by the theory that anxiety is a cause of good reading attainment. Such a theory would account for some of the apparently unconnected findings, especially: (*a*) the superior reading of girls might be due to their higher general level of anxiety, as reported by e.g., Burt (1937, p. 336) and Isaacs (1932); (*b*) the high verbal ability of anxious and obsessional neurotics might be due to their anxiety,; and (*c*) the poor verbal ability of the psychopathic personality might be attributed to the low level of anxiety, reported by e.g., Bowlby (1951) and Bender (1950).

It is not altogether easy to see why anxiety should manifest itself in this way. Perhaps the simplest explanation would be that the highly anxious child devotes a great deal of time to solitary reading, while the unanxious child is engaged in some more active pursuit. A similar explanation to account for sex differences in reading attainment has been put forward by Burt (1927, p. 196), who draws attention to the sedentary habits of the little girl, sitting at home reading while her brothers are out at play. This sex difference may be due to a more fundamental temperamental difference. Many writers have noticed that the anxious, solitary child may become an avid reader to satisfy his needs in fantasy, as an escape from the real world or simply from the ambition for success which often characterises anxious children. It is less likely that anxious children would get pleasure out of doing arithmetic, and it is probable that anxiety has a disorganising effect on arithmetic, as Burt (1937) observed. There is considerable experimental evidence that anxiety impedes the acquisition of complex skills (e.g., Gordon and Berlyne, 1954) and arithmetic, for the school child, is undoubtedly a complex skill.

(4) The results suggest several hypotheses for further research. (i) If reading-arithmetic disparity and high VEd : low KM are both part of the same general syndrome, it would be expected (*a*) that psychopathic personalities would have poor reading attainment in relation tc arithmetic ; (*b*) that anxiety would bear a similar relationship to high VEd : low KM as it appears to have (according to the present findings) to reading-arithmetic disparity. In this connection it is interesting to note that girls tend to have high VEd : low KM (Vernon, 1950b, p. 32), and this might be explicable in terms of their higher average level of anxiety.

(ii) If anxiety is induced and developed, even in part, by patterns of parental discipline and behaviour, then there should be some relationship between these patterns and reading-arithmetic disparity and intelligence sub-test scatter. Evidence supporting this hypothesis does in fact exist (Levy, 1943 ; Feinberg, 1949 ; Davis and Kent, 1955), but consideration of its interpretation would take us too far field from the present paper.

ACKNOWLEDGEMENTS.—I wish to acknowledge my debt for the advice and encouragement of Dr. D. Russell Davis and Professor O. L. Zangwill of the University of Cambridge, where this research was carried out ; to the friendly co-operation of the schools in which the research was carried out ; and to King's College, Cambridge, and the Pinsent-Darwin fund for their financial support.

V.—REFERENCES.

Altus, W. S., and Clark, J. H. (1949). The effect of adjustment patterns upon the intercorrelation of intelligence sub-test variables. *J. Soc. Psychol.*, 30, 39-48.

Ayres, L. P. (1909). *Laggards in Our School.* New York : Russell Sage Foundation.

Bender, L. (1950). Anxiety in disturbed children. In *Anxiety,* ed. Hock, G. H., and Zubin, J. New York : Grune and Stratton.

Bowlby, J. (1951). *Maternal Care and Mental Health.* Geneva : Bull. World Health Organisation.

Burt, C. (1937). *The Backward Child.* London : Univ. London Press.

Chase, V. (1932). Educational achievement of delinquent boys. *J. Juv. Res.*, 16, 189-192.

Davis, D., Russell, and Kent, N. (1955). Intellectual development in school children, with special reference to family background. *Proc. Roy. Soc. Med.*, 48, 993-995.

Eysenck, H. J. (1947). *Dimensions of Personality.* London : Kegan Paul.

Feinberg, H., and Reed, C. L. (1940) : Reading level of a group of socially maladjusted boys. *J. Soc. Psychol.*, 12, 31-38.

Feinberg, H. (1947) : Achievement of a group of socially maladjusted boys as revealed by the Stanford Achievement Test. *J. Soc. Psychol.*, 26, 203-212.

Feinberg, H. (1949). Achievement of a group of children in foster homes, as revealsed by the Stanford Achievement Test. *J. Genet. Psychol.*, 75, 293-303.

Feinberg, H., and Moscovitch, E. (1956) : Achievement on the Stanford Achievement Test of children placed in difficult own home situations compared with children placed out of the home. *J. Exp. Educ.* (In press.)

Fendrick, P., and Bond, G. (1936). Delinquency and reading. *J. Genet. Psychol.*, 48, 236-243.

Hallgren, B. (1950). Specific dyslexia (congenital word-blindness) : A clinical and genetic study. *Acta. Psychol. Kbh. Suppl.*, 65, xiii, 287.

Himmelweit, H. T. (1945). The intelligence-vocabulary ratio as a measure of personality. *J. Person*, 14, 93-105.

HIMMELWEIT, H. T., and PETRIE, A. (1951). The measurement of personality in children. *Brit. J. Educ. Psychol.*, 21, 9-29.

JASKAK, J. (1941). *Wide Range Achievement Test.* Wilmington: Story.

ISAACS, S. (1932). Neurotic difficulties in young children. *Brit. J. Educ. Psychol.*, 2, 257-265.

LEVY, D. M. (1943). *Maternal Over-protection.* New York: Columbia Univ. Press.

LINCOLN, E. A. (1927). *Sex Difference in the Growth of American School Children.* Baltimore.

LYNN, R. (1955). Personality factors in reading achievement. *Proc. Roy. Soc. Med.*, 48, 996-998.

LYNN, R. *The Relation of Anxiety to Educational Attainment in School Children.* Cambridge University: Unpublished Ph.D. thesis.

PIQUER Y JOVER, J. I. (1946). *El nino atardonado y delincuente.* Madrid.

SCHONELL, F. J. (1942). *Backwardness in the Basic Subjects.* Edinburgh: Oliver and Boyd.

STROUD, J. B., and LINDQUIST, E. F. (1942). Sex differences in Achievement in the elementary and secondary schools. *J. Educ. Psychol.*, 33, 657-667.

TAYLOR, J. A. (1953). A personality scale of manifest anxiety. *J. Abnorm. Soc. Psychol.*, 48, 285-290.

VERNON, P. E. (1950). *In Reading Ability.* London: H.M.S.O.

VERNON, P. E. (1950). *The Structure of Human Abilities.* London: Methuen.

WECHSLER, D. (1941). *The Measurement of Intelligence.* Baltimore: Williams and Wilkins.

(*Manuscript received, March 7th, 1956.*)

SUMMARIES OF RESEARCHES REPORTED IN DEGREE THESES[1]

Educational Provision for Able Children not Selected for Grammar or Technical Schools.

By ALAN MOON

(Summary of thesis accepted (1955) *in part qualification for the degree of M.A. in Education in the University of Birmingham.)*

THE AIM OF THE INVESTIGATION.

There is a wide diversity of secondary school courses in England, and it was hoped that an examination of some of them would provide useful information relevant to secondary guidance at 11+. Particularly, it was hoped to gain information about the variety and types of courses suitable to the 11+ borderline group.

A SURVEY OF SECONDARY EDUCATION IN 1952.

The 145 counties and county boroughs in England and Wales were asked for information about secondary provision other than that in grammar and technical schools. Useful replies from 130 revealed the following types :

(i) Unreorganised all-age schools.
(ii) Non-selective modern schools, where a supplementary inquiry of a limited nature showed that some children in a few schools were taking G.C.E., although only a very few in any one school.
(iii) Comprehensive schools.
(iv) Multilateral schools.
(v) Selective (central) schools, filled after the grammar schools. Information from twenty-four such schools showed that in most of them, suitable children were encouraged to take G.C.E. at ordinary level, thus suggesting that such schools catered for the marginal ' academic ' pupils.
(vi) Modern schools providing special vocational courses, the commonest being technical courses for boys, pre-nursing courses for girls, and commercial courses, including shorthand and typewriting. Such courses, filled from children allocated to their schools at 11+, were reported by fourteen authorities. Another eight authorities reported reallocation to such courses from other modern schools after 11+, thus giving pupils a wider choice of bias. It appeared that vocationally biased courses could be made vehicles of general education attracting children to stay at school until 16+.

Special investigations were now made with limited groups of children in the last two types of schools above.

ANALYSIS OF 11+ DATA CONNECTED WITH BORDERLINE G.C.E. CANDIDATES.

A central school was chosen receiving a mixed intake of 12 per cent. of 11+ population of a county borough after the grammar schools had taken 20 per cent. to 21 per cent. About one-third stayed for a fifth year to take G.C.E. at ordinary level, and inquiry suggested that most candidates suited by interests and attitudes stayed. The group numbered 245.

Mean I.Q. Score = 107·2. Mean Arithmetic Score = 106·2
Mean English Score = 107·0 Mean Total Score (with two intelligence tests) = 427·7.

Comparisons were made between pupils in the group using simple analysis of variance. Firstly the hypothesis that it was not possible to distinguish at 11+ between pupils who subsequently proved suitable for G.C.E. and those who did not, was tested. The total group was divided into eight, by sex, year of entry (1948 and 1949) and G.C.E. candidature. Comparisons by Total Scores, Intelligence Scores and English Scores showed no significant F-Ratio. Arithmetic Scores were significant at the 1 per cent. level, but splitting the total group into two (G.C.E. and Non-G.C.E.) gave no significant t- statistic with arithmetic scores. It appeared that suitability for a G.C.E. course could not readily be determined within the central school group at 11+.

---

[1] Theses Outlines must be submitted through the Head of the Department in which the research was carried out.

The second hypothesis tested by analysis of variance was the impossibility of distinguishing at 11+ between better and poorer performers in G.C.E. The seventy pupils taking the examination were given a subject score (0 to 7 subjects passed). One case only scored seven and was omitted.

Analysis of variance failed to distinguish between the remaining seven groups either by Total, Intelligence, Arithmetic or English scores, thus leaving the null hypothesis undisturbed.

Similar data from Moray House tests at 11+ were now obtained for a grammar school group selected as a borderline group by another authority:

No. in Group=74. Mean I.Q. score =118·4
Mean Total score =456·6.

Comparison of Total Scores with the central school group (t-test) was highly significant, suggesting that the new group was more able at 11+. The G.C.E. subject scores for both groups were compared by t- tests (once ignoring non-examination pupils in both groups, and once giving them zero scores), and neither proved significant. Roughly one-third took G.C.E. in each group.

This suggested the difficulty of deciding suitability for G.C.E. among 11+ borderliners and the importance of interests and attitudes. It seemed that possible G.C.E. candidates might score below the usual level of grammar school entry at 11+.

COMPARISONS BETWEEN THE CENTRAL SCHOOL GROUP AND CHILDREN IN VOCATIONALLY BIASED COURSES.

Suitable Moray House 11+ scores were obtained to enable Total Scores to be calculated for fourteen children in a commercial group, sixteen in a pre-nursing group, thirteen in an engineering group, and twelve in a retail group (girls preparing for shop work).

T-test comparisons by total scores with the central school group showed no significant differences. This suggested that these courses might be suitable alternatives to G.C.E. for children of ' central school group ' calibre.

SUMMARY OF CONCLUSIONS.

Statistical evidence suggests that:

1.—There are G.C.E. candidates below the normal level of grammar school selection (i.e., in the central school group with I.Qs. ranging roughly between 100 and 118 or 120).

2.—It is difficult to distinguish G.C.E. and non-G.C.E. candidates at 11+ among borderliners or to predict subsequent G.C.E. performance.

3.—Factors other than 11+ ability and attainment are important and they may include interests and attitudes.

4.—Educational guidance into various courses (possibly including inter-school movement) needs to continue after 11+.

5.—A range of courses (vocationally biased and G.C.E.) may be needed for this group of able children, as well perhaps for lower grammar school streams. It is tentatively suggested that the proportion should be : vocational—2 to G.C.E.—1.

While the statistical analyses are based on small groups, the earlier survey suggests they may be of some value to those concerned with guidance in secondary schools.

BOOK REVIEWS

ASSOCIATION OF TEACHERS IN COLLEGES AND DEPARTMENTS OF EDUCATION (1956). *The Supply of Mathematics and Science Teachers.* London: Methuen, pp. 56, 3s. 6d.

H. S. DYER, R. KALIN, and F. M. LORD (1956). *Problems in Mathematical Education.* Princeton, New Jersey: Educational Testing Service, pp. 50.

The American survey is the summary of an investigation carried out by the Educational Testing Service in 1955. The reasons for the shortage of mathematics teachers in schools and colleges are examined under three main headings: the learner, the teacher and the curriculum. Many bright youngsters abandon mathematics as soon as possible and this subject has the " dubious honour of being the least popular in the curriculum." Mention is made of psychological theories—the associationist, field theories and the work of Piaget; but no particular method of teaching is advocated because, although much work has been done on the psychology of reading and language, there is no comparable work on the psychological principles that govern the learning of mathematics. Co-ordinated basic research into the processes of learning mathematics must replace piecemeal research.

The attitude of the teacher to mathematics is a vital matter; many teachers have a long standing distaste for mathematics and many " are ignorant of the mathematical basis of arithmetic." Students are accepted for teacher training with very little knowledge of mathematics and few training colleges provide adequate method courses. Solutions to this problem are difficult to find; greater attention to the minimum requirements for professional certification might help, but a more direct attack on the problem might be made by holding courses in remedial mathematics and by reducing the size of classes. Classes of more than thirty-five pupils with a wide range of arithmetical ability are blamed for the present attitude to arithmetic, and a fundamental change in the conditions of teaching is absolutely necessary.

The curriculum has not changed basically over the last century; it has failed to grow with the times and has never been presented to the pupil in a way that seems reasonable and natural. Pupils too often " wrestle with mathematical ideas before they have grown enough to handle them successfully " and " the bright and the dull are kept busy with work which lacks any challenge or satisfaction." The American report presses for a change in popular attitude towards the subject, if only because in this technological age " a command of numbers can be just as important as a command of words." The survey ends with a full and useful bibliography with 250 references to research on teacher training, curriculum and the psychological theories of mathematical teaching. Unfortunately, there are few references to similar work over here, e.g., by Nunn, Hamley and Schonell, etc., and the only reference to Burt's work is the 1928 L.C.C. report on Arithmetic.

The A.T.C.D.E. report sets out to be an accurate account of the position of mathematics and science in the training of teachers. Its arguments are supported by statistics from the Ministry of Education, training colleges and from the results of a questionnaire sent to 101 training colleges. Like its American counterpart, this report stresses the distaste for mathematics started in the primary schools by teachers who themselves have little interest in its teaching. The lack of balance between school and training college time tables and the frequent lack of any mathematical qualification among training college entrants, all contribute to the present crucial situation.

Many two-year trained teachers go on to grammar schools to teach Mathematics and Science, but this is poor consolation to the secondary modern schools. Again, concern at the low standard of attainment in mathematics has frequently been commented upon (Cf. 'The mathematical attainments of students entering training colleges,' S. A. Wallbank, *Researches and Studies*, No. 11, January, 1955). This report realises that part of the solution must depend on the quality of student entry to training colleges and the provision of adequately trained (and motivated) mathematics and science staff. There should be at least one mathematics graduate and one science graduate on the staff of every training college. Applicants to training colleges should, if they show necessary ability, be urged to study and to teach mathematics or science.

Both reports spring from similar sets of circumstances ; both have no simple solution to offer. The Americans stress the need for curriculum reform and comprehensive psychological research, and the A.T.C.D.E. look to the training colleges and Area Training Organisations for better trained mathematics teachers. W.H.K.

L. A. AVERILL. *The Psychology of the Elementary School Child.* Longmans, Green and Co., pp. 459, 30s.

This is a readable book for the elementary student, but it is superficial in treatment, and at important points fails to get down to fundamentals. It is apt also to be somewhat verbose, and as so often with American books, pays little attention to important work in British researches on childhood and education. C.W.V.

A. M. B. BLACKWELL (1956). *Lists of Researches in Education and Educational Psychology, Supplement II.* Nat. Found. Educ. Res. Publ. No. 9. London : Newnes, pp. 62, 15s.

Once again Mrs. Blackwell and the National Foundation have put us in their debt by bringing up to date their lists of theses. These are quite invaluable to research students in Education Departments, and to those who advise such students on their choice of subjects. It is unfortunate that considerations of cost make it necessary to issue two-yearly supplements ; we now have to search through four volumes, and there is not even a cumulative index. I have calculated that the list expands at the rate of about 6 per cent., or eighteen pages per year. Could not the next publication (in 1958) consist of an amalgamated edition of some 210 pages, printed on one side of the paper only, so that further additions—say for the following ten years, could be stuck onto the blank pages ? P.E.V.

MARGARET MEAD and MARTHA WOLFENSTEIN (edit.) (1955). *Childhood in Contemporary Cultures.* University of Chicago Press, pp. xi+473, 59s. 6d.

The fourteen contributors to this collection of articles include anthropologists, psychologists and psycho-analysts and their papers fall into three groups ; observational studies of children, analyses of material made by, about or for children and accounts of various kinds of interview.

The observational studies deal with various aspects of child behaviour in widely different cultures such as in Bali, France and a Jewish community in Eastern Europe. Each covers a limited sample of child behaviour, e.g., music and art in Bali, or play in a French park, and is chosen to reflect some major cultural characteristic of the children's social environment.

Of the second group, the first articles compare child rearing literature in America, Germany and Russia. They show the great change in American thinking about children which has taken place recently, and that present-day Russian and German ideas tend to resemble those prevalent in America 100 years ago. These are followed by studies of children's stories, contemporary films and children's imaginative productions to show how they are related to many complicated factors in adults' relations to children. An exhibition of paintings by French children and the results of a projection test given to young Germans are analysed so as to reveal the underlying cultural characteristics.

In the third group a number of admittedly highly speculative interpretations, based on psycho-analytic theory, are given to the results of interviews with parents and children. Finally, the editors try to indicate some of the implications, for child rearing practice, of the insights gained from such studies as are outlined here and the difficulties in the way of their application.

Most of the articles make interesting reading and, by their grouping, the editors have made an unconnected series of papers into a coherent whole. It is probable that readers of this *Journal* will find the descriptive work most rewarding, for this indicates lines of enquiry which may be fruitfully explored, expecially in the production of tentative hypotheses which may, later, be tested by experiment. The active worker in the field will appreciate the extensive bibliographies and references which follow each article. L.B.B.

NATIONAL ASSOCIATION FOR MENTAL HEALTH (1955). *Periods of Stress in the Primary School.* London: N.A.M.H., pp. ii+38, 3s.

This is a little book on a very large subject; it surveys the possible peak periods of stress in the primary school—entry; commencement of learning; class and school transfers. It also comments on general sources of stress—family relationships; staff relationships and poor amenities.

The book is clearly stated to be a summary of the opinions of experienced field workers, suggesting areas of research and raising issues for discussion. Modern trends in educational thought have obviously influenced their opinions as well as everyday experience. Activity methods, democratic type discipline and good personal relationships in the classroom are valued as tending towards better mental health and less stress.

In making this book brief and to the point it has, perhaps unfortunately, been necessary to give the same emphasis to opinion based on practical experience as to the evidence of research. In dealing with 11+ perhaps the book is not quite consistent with its intention of suggesting research, although it would certainly provoke discussion; this section is written as if the issue were in no doubt at all, and one may wonder whether the situation is as simple as portrayed and the 11+ examination the real *bete noire.*

There are two main conclusions—the need for research and the potential effect of schools on mental health. The first is well known but needs to be kept well in mind when reading the text. The second needs to be given much more consideration amongst educationists and this book makes an excellent contribution in bringing it to the attention of the reader. H.J.W.

R. PEDLEY (1956). *Comprehensive Education: A New Approach.* London: Gollancz, pp. 203, 13s. 6d.

Unexpectedly, Mr. Pedley no longer supports the comprehensive secondary school, except as an intermediate or common school for a few hundred boys and girls aged 9 to 15. He dislikes the current monsters for two thousand odd pupils (especially when non-coeducational), partly because it is difficult for pupils and staffs to develop good social relations and a sense of security, but partly because—even with their numbers—it is impossible to provide sufficient specialist staff to teach a wide variety of subjects to reasonable-sized groups of sixth-form pupils. His solution, therefore, is the county college from 15 to 19, both for part-timers, and full-time technical students and academic streams. He seems to make a good case on educational grounds, but the book is noteworthy too for its concern with the psychology of the adolescent—the grammar school pupil who resents his childish status, and the incipient Teddy Boys or sexually mature girl who sees no point in the sort of education most modern schools provide. The counter-argument which is sure to be raised—the value for character-training of the prefect system in a school for 11 or 13 to 19 year olds—is rejected by him as an outworn shibboleth.

Many readers may, perhaps, be put off by his first chapter, which presents all the usual left-wing criticisms of 11+ selection, in the usual tendentious fashion. Because the present reviewer has stated that the ability measured by intelligence tests is largely acquired, that does not mean that it is not also remarkably stable in the majority of pupils from 11 to 15 or 18. And the existence of a tiny percentage of modern school pupils who get good G.C.E.s does not prove that " the process of selection is riddled with weaknesses at every stage . . . its instruments unscientific and its standard of successful diagnosis low." However, if read with a grain of salt, there is also much good sense and an imposing collection of statistics on the need for radical reforms in secondary education. P.E.V.

A. A. ROBACK (edit.). *Present-day Psychology.* New York: Philosophical Library, pp. 995, $12.00.

The basic idea which governed the Editor's planning of this book was that far too large a proportion of psychologists are narrow specialists and have had little training in general psychology, and that they ought to be familiar with the broad issues and main results in other branches of psychology. With that general principle the present reviewer is in strong agreement. Far too many psychologists begin research on a problem before they have sufficient background to see their own problem in its true relationship to others.

Professor Roback has taken a great deal of trouble in preparing this book. He reports that it involved writing seven hundred letters, and very frequently he could not get the writer that he would have chosen for the particular topic.

In its final form we have forty chapters by nearly as many separate authors dealing with practically every aspect of psychology. Unfortunately, however, it has proved impossible to arrange that the various chapters shall take a similar form. Sometimes they are very broad indications of general trends ; at other times they are much more factual with specific references to authors. Further, the quality of the contributions varies enormously ; thus, while most students early in their course would profit greatly from reading, say, Guilford's chapter on ' Psychometry,' there are several other chapters which would themselves need an expert in the particular branch to see what the author was trying to say, since the outlines are so arid and general.

Among the authors I can only find one from Great Britain, Professor Pickford. I hope it will not be taken as national prejudice if I say that at least his chapter on ' Factorial Studies of Aesthetic Judgments " is better than some of the others.

I doubt whether the volume is worth including in the library of any college or department primarily concerned with educational psychology, especially in view of its cost. On the general editing two comments should be made. First, there is no subject index ; second, in a book of this type cross-references should be facilitated as much as possible, but the heading of every one of the 995 pages is " Present-day Psychology." It would have been much more useful if the name of the author of each chapter had been on one side, and part of the title of the chapter on the other.

C.W.V.

LELAND H. STOTT (1955). *The Longitudinal Study of Individual Development,* Detroit: The Merrill-Palmer School, pp. x+115, $2.75.

Dr. Stott's short guide to the longitudinal study of child development is based on the methods of individual case study which are used at the Merrill-Palmer School, Detroit, where each student makes a detailed study of one child. His aim here, is therefore essentially pedagogic.

There can be little doubt that a student using the methods described would learn a great deal about assessing growth and development, and would gain a useful insight into the life of the child he studied. Thus, starting with height and weight, and a consideration of growth curves, the student is offered, amongst other things, ways of estimating the maturity of eating behaviour and sphincter control, speech, social and emotional behaviour, personality and intelligence. Hardly surprisingly, the attempt to express the development of these functions in terms of curves meets with most success in the case of physical growth, and becomes more tenuous in the case of emotional or social behaviour. In the latter cases one wonders just how far the author's method of arranging a series of " maturity indicators " in a developmental sequence could be rendered more precise, in view of the many variations which children display.

At the back of the book is a large and interesting collection of curves, profiles and charts for displaying graphically the information obtained, but the author is under no illusion that all the relevant information can be charted, and draws attention to the importance of writing notes of observations, and impressions.

In regard to the family, a useful list of qualities for rating includes the adequacy with which mother and father fulfil their respective roles ; the emotional climate of the home ; and the kind of parental guidance exercised. Many of these ratings, as the author confesses, are of necessity made on the basis of rather sketchy evidence, and the reliability and validity of some of the scales are not mentioned. Methods of handling socio-economic data are also provided.

A major criticism that I would make is that there is no reference to the subjective aspects of development : the child's phantasies, play themes, modes of thought, his perceptions of his family and his surroundings, or of his purposes and values. This remains the weakness of a great deal of American work on child development, and, though difficult to accomplish, one would like to see some attempt to treat of such matters in a guide of this kind. C.B.H.

PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

J. P. GUILFORD (1956). *Fundamental Statistics in Psychology and Education* (3rd edit.) McGraw-Hill, pp. xi+565, 47s.

B. KLOPFER, *et al* (1956). *Developments in the Rorschach Technique*. Vol. 11. Harrap pp. xx+828, [illegible]0s.

C. T. MORGAN (1956). *Introduction to Psychology*. McGraw-Hill, pp. xviii+676, 45s.

✓ J. M. STEPHENS (1951). *Educational Psychology*. New York : Holt, pp. xxiii+692, $6·00.

F. BAUMGARTEN (1956). *Psicologia E relazioni Umane Nell'Azienda* (trans. by L. Rocco from *Die Psychologie der Menschenbehandlung im Betriebe*). Turin : Societa Editrice Internazionale, pp. xxiv+287, L.1200.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (1956). *L'Analyse Factorielle, et ses Applications*. Paris, pp. 431, Fr. 1,500.

P. CHAUCHARD (1956). *Le Langage et La Pensée*. Paris : Presses Universitaires de France, pp. 128. No price given.

M. MEGRET (1956). *La Guerre Psychologique*. Paris : Presses Universitaires de France, pp. 128. No price given.

L. WAY (1956). *Alfred Adler : An Introduction to His Psychology*. Penguin Books, Ltd., pp. 252, 3s. 6d.

J. C. MORRISON (edit.) (1956). *Who are the Puerto Rican Pupils in the New York City Public Schools ?* New York : Board of Education, pp. xv+88.

M. PRUDHOMMEAU (1956). *Les Enfants Déficients Intellectuels*. Paris : Presses Universitaires de France, pp. vii+305, Fr. 1,200.

✓ N. MUKERJI (1954). *An Enquiry about Examination*. Calcutta : Das Gupta, pp. ii+67, Rs. 2/8/-.

C. CHERRY (edit.) (1956). *Information Theory : Third London Symposium*, 1955. Butterworth, pp. xii+401, 70s.

✓ R. PEDLEY (1956). *Comprehensive Education : A New Approach*. Gollancz, pp. 203, 13s. 6d.

✓ S. BRAHMACHARI (1954). *Social and Moral Attitudes*. Calcutta : Orient Longmans, pp. 113, Rs. 5.

B. SPOCK (1951). *The Pocket Book of Baby and Child Care*. New York : Pocket Books, Inc., pp. ix+502, 4s. 6d.

P. CHAUCHARD (1956). *La Maitrise du Comportement*. Paris : Presses Universitaires de France, pp. 224, Fr. 800.

✓A. M. Blackwell (1956). *Lists of Researches in Education and Educational Psychology, Supplement II.* National Foundation for Educational Research Publication No. 9. Newnes, pp. 62, 15s.

H. S. Lippman (1956). *Treatment of the Child in Emotional Conflict.* McGraw-Hill, pp. x+298, 45s.

S. N. Eisenstadt (1956). *From Generation to Generation.* Routledge and Kegan Paul, pp. 357, 42s.

A. M. Rose (edit.) (1956). *Mental Health and Mental Disorder.* Routledge and Kegan Paul, pp. xii+626, 40s.

F. E. Schonell (1956). *Educating Spastic Children.* Edinburgh: Oliver and Boyd, pp. xv+242, 21s.

E. B. Hurlock (1956). *Child Development* (3rd edit.). McGraw-Hill, pp. xvi+703, 45s.

G. Ambrose (1956). *Hypnotherapy with Children.* Staples, pp. 135, 10s. 6d.

P. Osterrieth, J. Piaget, *et al.* (1956). *Le Probléme des Stades en Psychologie de L'Enfant.* Paris: Presses Universitaires de France, pp. 163, Fr. 700.

S. N. Mehrotra (1956). *A Group Guidance Project.* Bureau of Psychology, U P., Allahabad (Publication No. 7), pp. iv+51. No price given.

F. Baumgarten (1956). *Berufs- und sozialpsychologische Untersuchungen bei Arbeitern.* Zürich: Rascher Verlag, pp. 150, Swiss Fr. 16.30.

B. S. Bloom, *et al.* (1956). *Taxomomy of Educational Objectives, Handbook I: Cognitive Domain.* Longmans, Green, pp. 207, 10s. 6d.

D. H. Stott (1956). *Unsettled Children and their Families.* University of London Press, pp. 240, 16s.

H. S. Broudy and E. L. Freel (1956). *Psychology for General Education.* Longmans, Green, pp. viii+456, 35s.

P. D. Leedy (1956). *Reading Improvement for Adults.* McGraw-Hill, pp. viii+456, 34s.

V. Zuckerkandl (1956). *Sound and Symbol: Music and the External World.* Routledge and Kegan Paul, pp. vii+399, 32s.

V. Mallinson (1956). *None Can Be Called Deformed.* Heinemann, pp. x+214, 12s. 6d.

E. Eaglesham (1956). *From School Board to Local Authority.* Routledge and Kegan Paul, pp. ix+220.

L. F. Shaffer and E. J. Shoben (1956). *The Psychology of Adjustment* (2nd edit.). Boston: Houghton Mifflin; London: Constable, pp. xxi+672, 42s.

REQUEST FOR BACK NUMBERS

The Graduate School of Education of the new bilingual International Christian University of Japan seeks gifts of back volumes, or of individual numbers, of the *British Journal of Educational Psychology*, except for Vols. 25 and following which are already in its library. Please write to Miss Tane Takahashi, Librarian, 1500 Osawa, Mitakas-hi, Tokyo, Japan, listing what volumes or numbers you are able to offer, and she will check possible duplications and send directions for dispatch.

TAVISTOCK

DISCUSSIONS ON CHILD DEVELOPMENT (*Vols. I & II*)
J. M. Tanner and Bärbel Inhelder (*Editors*)

These volumes comprise the carefully edited transcripts of the discussions at the first and second meetings of the World Health Organization Study Group on the Psychobiological Development of the Child.
Demy 8*vo, Vol. I*, 25*s. net* ; *Vol. II* (*illustrated*), 28*s. net*

THREE TITLES BY THE DISTINGUISHED PSYCHO-ANALYST AND PAEDIATRICIAN

D. W. Winnicott

THE CHILD AND THE FAMILY: First Relationships

Primarily addressed to parents of young children, this book will also be found of absorbing interest to teachers, social workers, and psychologists—indeed to all who are professionally concerned with the care of children.
Demy 8*vo*, 12*s*. 6*d. net*

THE CHILD AND THE OUTSIDE WORLD:
Studies in Developing Relationships

This work contains writings on the theory and practice of child care, with emphasis on relationships with the outside world of the growing child and adolescent. *Demy* 8*vo*, 16*s. net*

COLLECTED PAPERS: Through Paediatrics to Psycho-Analysis

This work collects together Dr. Winnicott's distinctive and varied contributions addressed to scientific audiences. They are offered for the consideration of a wide readership including all whose work or practice involves the care of children in health or sickness.
Medium 8*vo*, *about* 35*s. net*

EMOTIONAL PROBLEMS OF EARLY CHILDHOOD
Gerald Caplan (*Editor*)

An authoritative statement by a group of leading specialists on the diagnosis, treatment, and prevention of emotional disturbance. " This is a book that will give ample food for thought to both specialist and layman."—*Nature*.

" This book contains much that will be of interest to psychiatrists and psychologists."—*British Medical Journal*.
Medium 8*vo*, 42*s. net*

TEACHER, PUPIL, AND TASK:
Elements of Social Psychology Applied to Education
Oscar A. Oeser (*Editor*)

" We should therefore be grateful to Professor Oeser and his colleagues for bringing the educational applications of social psychology to our notice in a small compass and in a readable form. A book of this kind is most welcome and was definitely needed."—*A.M.A. Journal*.

" On the whole Professor Oeser has succeeded in producing the most useful and comprehensive treatment of the topic that has been published to date."—*The Times Educational Supplement*.

" On all counts this is a book to buy."—*Higher Education Journal*.
Demy 8*vo*, 18*s. net*

PUBLICATIONS

VOLUME XXVII. JUNE 1957 PART 2

SYMPOSIUM: THE DEVELOPMENT OF CHILDREN'S MORAL VALUES

I.—SOME ASPECTS OF MORAL DEVELOPMENT IN A CHANGING SOCIETY

By JAMES HEMMING

MR. HEMMING was requested to contribute the following article as a general introduction to a symposium on the theme of children's moral development as an object of scientific study. Subsequent contributors will deal with it from diverse standpoints. It should be noted that 'moral' should not be taken to mean 'conforming to the moral standards conventional in our own society,' but more broadly as referring to the 'internalised controls' or 'values' that guide people's conduct.—EDITOR.

I.—INTRODUCTION.

SOCIAL scientists have been inclined, until quite recently, to leave the specific study of moral development outside their field of inquiry; the confusion and prejudice with which the word 'moral' was surrounded hardly encouraged them to do otherwise. This withdrawal from active concern with moral issues, however, has gradually become untenable. The social anthropologists have shown that man and his culture must be studied in relation to each other. A moral structure always forms part of a culture pattern; thus, psychologists have been drawn into studying moral development and its relationship to the growth process. Moreover, in the years following the war, character disorders have become an increasingly important category of psychological illness. On practical as well as theoretical grounds, therefore, the study of moral development as an aspect of growth claims scientific investigation. So much is obvious, but what is not so clear is how this development may most usefully be studied and how such a concept as 'moral maturity' should be defined. Neither can be dealt with here except in the most limited and tentative way. When the potential field of inquiry is so complex, the most one can hope to do is to dig a little in a small area of it.

At its simplest, the moral development of the child may be described in terms of the process by which the child acquires the values esteemed by his community, how he acquires a sense of right and wrong in terms of those values, and how he learns to regulate his personal desires and compulsions so that, when a situational conflict arises, he does what he ought to do rather than what he wants to do. In brief, to look at it the other way round, moral development is the process by which a community seeks to transform the egocentricity of the baby into the social behaviour of the mature adult. An obvious role of psychology in studying moral development is to describe as accurately as possible the details of the processes by which values and the capacity for self-regulation are acquired and, conversely, to describe the conditions and deprivations which inhibit or distort moral growth.

However, psychologists have to probe beyond the simple transfer of value systems from one generation to another if they are to offer, eventually, a reasonably complete description of moral growth and moral maturity in the modern world. This for two reasons. In the first place, the socially tidy situation when a child's mode of moral development was to acquire for himself a set of standards and values clearly manifest in the culture around him has now disappeared. Almost everywhere in the world conflict, not clarity, is becoming the general rule. Major dichotomies split established social outlooks—materialism *v.* Christianity ; Roman Catholicism *v.* atheism ; communism *v.* Confucianism—bringing many minor fragmentations of traditional value systems in their train.

Secondly, the process of simple cultural transmission of values is today complicated by the comparatively frequent influx of new knowledge which challenges moral preconceptions. This means that no code of values extant during a child's early years will remain completely unaffected during the period while the child is growing up. For example, the quite recently acquired understanding—by the community at large—of environmental and deep psychological influences as factors in criminal behaviour, undermines, without necessarily *removing*, moral attitudes which see the complete source of delinquency and crime in the malevolence of the individuals concerned. Or, again, knowledge about the rate of increase of world population, only abstracted with any accuracy during the past ten years, challenges moral preconceptions about birth control. New knowledge alters the human situation, and the function of morals is to control the human situation as apprehended in any culture ; hence an inevitable relationship exists between new knowledge and moral change.

In a changing society, in fact, the habit of conformity is not enough. Once culturally standardized answers to any major ethical questions cease to exist, an element of selection becomes part of moral behaviour ; and the capacity to select with judgment becomes a feature of moral maturity. When existing values are likely to be tested again and again by the extension of knowledge, the capacity to *re-assess* values in the light of fresh evidence also becomes a necessary attribute of the morally mature person. Indeed, perhaps the most pertinent question a psychologist can ask himself in the field under discussion is what the conditions are which permit a ' schema of values ' to be laid down in the growing mind while preserving flexibility of outlook.

In this paper, an attempt will be made to relate together principles covering the satisfactory absorption of socially esteemed values during childhood and the development of those necessary powers of selection and flexibility upon which moral maturity in a changing society depends.

II.—The process of acquiring values.

Just as a physicist finds it convenient to start with the concept of a theoretically simple system and advances by studying observed modifications of the system, so a psychologist may usefully begin the study of moral development in a complex society from the process of value acquisition as it appears to occur in an uncomplicated primitive society. " There is no doubt," writes Margaret Read, " that in relatively static societies which are clearly delimited, and in a cultural sense isolated, the process of socialisation can be most effectively studied."[1]

The combined findings of cultural anthropology, child psychology and clinical psychology provide us with a fairly complete theoretical picture of how values are acquired and introjected so that social conformity is assured. In a primitive society, the new-born child and his family group exist within a homogeneous culture of which the social, economic, legal and religious aspects interlock as a self-consistent pattern of values, beliefs, obligations and relationships.[2] From the moment the child becomes aware of the independent existence of others, he begins to absorb the culture through the mediation of his parents or parent substitutes, and, later, through contact with other members of his family group. Imitation, emulation and identification are all a part of this process of absorption. But, even in the most permissive culture, a child will not absorb the values which surround him without some conflict. He has to accommodate himself to the inconvenient facts that those who provide him with love, food and admiration can, and sometimes do, withhold them, and that other members of the family group sometimes enter into competition with him for what he wants for himself alone. Denial leads to frustration, anger, hatred—and fear of what hatred may bring upon him. This is the first moral crisis ; it is reduced to manageable proportions—to follow Freud—by the child's introjecting the values of society as he experiences them in the behaviour of his parents, and as he interprets them through his childish imagination ; thus, a rudimentary conscience is laid down.

From then on the child's process of moral growth is one of expanding involvement with his community, discovering his own individuality through interaction with others, and learning acceptable beliefs and social behaviour. Loving treatment from those who look after him, and his own need for approval, make the child eager to identify and accept just as, at appropriate stages of growth, he is eager to walk, talk and, in other ways, match up to the behaviour of the adult world around him. As the child grows, he takes on, one after another, the socially-expected rules of his age group together with the self-control they require ; he identifies with his peers, striving to emulate and excel in the approved patterns of behaviour and, simultaneously, he acquires a clearer perspective about himself as an individual. Each stage of conformity takes him ahead of his previous role and prepares him for the more significant role which is to follow ; initiation at puberty confirms his past advance and opens the way to new responsibilities. He may be a member of one sub-group or another within his community—an in-group of aristocrats, perhaps, or of warriors—but this does not upset the consistency of his moral milieu because the in-groups are not in conflict with each other but interlock and interact to form a unified culture. Whatever the age of a member of a primitive society, and whatever status he may attain within his community, all the forces of society—love and the withholding of love ; praise and ridicule ; reward and punishment ; the means to economic status ; success in marriage and fellowship ; the forms of ritual and ceremony, are concentrated upon the individual throughout his life to assure his conformity to traditional patterns of behaviour. Ultimately, if he attains the status of an elder, he identifies himself with the whole life of his community, past, present and to come, at the same time as he is most highly differentiated as an individual in his role of a person commanding respect because he exemplifies in a mature form the values most highly esteemed by the community. The elders set the tone and climate for the next generation and the pattern is perpetuated.

Malinowski has warned us against the naîve idea that the existence of an unambiguous pattern of culture makes it easy for all individuals composing it to behave well.[3] There are members of primitive communities who conform only grudgingly and some actual deviants ; but no member of the community is in doubt what *is* held to be right and what wrong by community standards, and a unified social pressure to accept and live by these standards is exerted ***consistently*** upon each generation of children.

As has already been noted, this "perfect circle" of culture transmission no longer exists except in a few very isolated communities. It is now the pattern of the past. Nevertheless, the simple, basic mechanisms of culture transfer still operate, although now overlaid by modifications in the conditions of transfer, and by the intrusion of greater variability generally.[4] We must now turn to consider the influence of these modifications and variations upon the child's moral development.

III.—THREE CONDITIONS OF EARLY MORAL GROWTH.

The processes of identification, self-differentiation and socialization—the essential elements of moral development—are unable, it seems, to go ahead smoothly in the absence of certain conditions. These cannot yet be described fully but three have now been established by a considerable weight of evidence.

(1) The work of Freud, Isaacs, Klein and others, has established that genuine, unstinted, undemanding parental love, or the love of parent substitutes, is the surest possible foundation for future moral maturity. Such love, as a welcome to the world, assures the child that life and people are 'good'—in so far as he is able to experience goodness through the satisfaction of his needs—and helps to lay down the attitudes which will later encourage him to move towards life and others, not away from them or against them.[5] Love, too, is the basis of the identification process. Denied objects for love, the child's emerging individuality lacks an anchorage, and growth on all planes is liable to be stunted. Lack of love also leaves the child a prey to too much anger and grief, training him in negative emotions at the expense of positive ones. Further, failure to form the right kind of identification relationships impairs the formation of a conscience: the result may be either a weak, inept conscience, responding unpredictably to moral challenges, or else an over-riding, demanding conscience which, by its rigidity and intolerance, acts as an obstacle to moral growth.[6]

(2) The second important condition appears to be the formation of a right valuation of the self in relation to others and in relation to life. Adler, Fromm and Horney have especially emphasized this. The early social environment speaks a language of its own to the child. It may delude him with the idea he is supreme; it may tell him he is worthless; it may speak with an uncertain voice; or it may assure him he is a likeable and appreciated person who can make a success of life. The relationships during the phase of toilet training appear to be particularly significant in this regard. Harshness and condemnation give a child an excessive idea of his own badness and, by contrast, make goodness seem something unattainable for him.[7]

Right personal evaluation is fundamental to sound moral growth because a person will only accept a responsible role in a situation in which he feels he has something to contribute. If discouragement about self-value is too great in the early years, a child may be left with no alternative but to contract out of social responsibility in favour of living by compensatory fictions.[8] Many workers have discovered that the delinquent—in spite of his bluff—is actually suffering under a deep sense of worthlessness.[9] An objective self-image, unconsciously incorporated in the psyche as the result of satisfactory early experiences, permits the emergence of a style of life appropriate to the individual and, therefore, capable of development; a distorted self-image, whether inflated or depressed, is likely to lead to false relationships, recurring discouragement, and compulsive compensations of an unsocial, if not actually anti-social, kind.

"If the parents deal many blows to the self-esteem of the child," writes Kluckhorn, "he can compensate by an exaggerated, defiant conformity to expectations, by accepting unimportance and dependence, by egotistical self-inflation. Different patterns of behaviour often represent the same underlying psychological cause. Aggressive and shy individuals may be just different outward manifestations of a wounded image of the self."[4]

(3) The third basic condition for moral growth in the early years seems to be sufficient elbow-room for the uniqueness of individuality to manifest itself. Jung and his co-workers go so far as to assume that there is within the psyche an innate movement towards the attainment of optimum development of personality. If this movement is too much impeded, repression results, leading, in its turn, to excessive, irrelevant responses to stimuli.[10] Fromm supports this view. "Destructiveness," he writes, "is the outcome of unlived life."[11] Lyward, who has worked with delinquents with conspicuous success, relates the maladjustment of many of his charges to a refusal on the part of parents to let their children be 'themselves.'[12]

Denied satisfaction of these three conditions, a child is liable to acquire a deep sense of grudge against society and to attack social values and conventions rather than to accept them. The role of the delinquent gang, for example, according to Cohen, is to "establish a set of status critieria in terms of which the boy can more easily succeed and (secondly) enable him to retaliate against the norms . . . by sanctioning aggression against those who exemplify and apply them."[13]

It will now be appropriate to consider the effects of other forms of personal and social variation upon moral development in a western culture.

IV.—VARIATION AND MORAL GROWTH.

A democratic society—particularly one in a state of rapid change—stimulates cultural variations by removing obstacles to the potentialities for innovation inherent in any community. Accordingly, the process of moral development in a modern democracy can only be understood if we take into account variations existing in the field of experience surrounding the growing child. We must also consider the relationship between personal variation and moral growth.

The main sources of personal variation are differences in intelligence and differences in temperament. Piaget has shown that the ability to make moral judgments matures gradually as the child grows up;[14] Gesell and his co-workers have worked out norms by age-groups for ethical ideas.[15] Individuals who fall below the norm will be, to that extent, retarded in their capacity for moral judgments and ideas and some will never attain mature moral insight. These will live their lives morally dependent, relying for moral stamina on what G. Allport calls "a moral backbone on the outside"[16]—the support of a clearly-defined structure of in-group mores. Temperamental differences are less well understood but their relationship to moral behaviour is established by much research into the origins of delinquency, crime and the psychopathic personality. Bowlby, for example, has demonstrated a relationship between delinquency and prolonged separation from "mothers or established foster-mothers during the first five years of life."[17] Many workers have shown broken homes to be a significant aetiological factor in delinquency. But all such research establishes, *pari passu*, evidence that a proportion of children remain robustly immune to extremely bad psychological circumstances. It may be argued that a more exact knowledge of the enrivonmental factors would explain why some children succumb and some do not in apparently the same circumstances; but the range of response, on existing evidence, is so wide that the assumption of considerable innate differences in children's resistance to environmental strains is unavoidable. Moral fibre appears to be a compound of what is inherited and what is experienced.

Environmental variations, in western society, are extremely wide and take many forms. There is a great range in the climate of love to which children are subjected; it runs from the stark rejection of an unwanted, illegitimate child at one extreme to the generous affection of an exceptionally happy and united family group at the other. Variations in the ideas of the parents on child training are equally diverse. Side by side in our community are homes where any kind of control is taboo and others in which authoritarian strictness survives. Class or neighbourhood differences produce systems of in-group mores that overlap in part but, in part, conflict sharply. On one side of the railway line respectability may be the criterion of social acceptance; on the other, a brush with the police is important for achieving group status.

Again, the pattern of values offered by the homes in our society varies from apathy and cynicism at one extreme to deep religious convictions at the other. Within a family group, different liked and respected authority figures may hold up different standards to the young.[18] There maybe disagreement on such fundamentals as the nature of man, the existence of God, the source of right and wrong, the validity of existing codes of sexual behaviour, and, indeed, on almost every ethical issue about which simpler communities offer, or used to offer, a single, well-integrated set of values.

A particular anomaly is the place which the code offered by the socially sanctioned religion takes in a changing society. In Britain the church is established, but its system of beliefs is accepted without question only by a minority; the Church has, in fact, become an in-group within the society it represents. Something of the traditional religion probably survives in most homes if in no clearer form than the parents' memories of their childhood, revived periodically by the Feasts of the Christian Year; but the content varies greatly. Even in the schools of England, where Religious Instruction is now compulsory, no single pattern of values is manifest and the internal value systems must often be confused. A recently unpublished inquiry into the ethical outlook of second-year students in two teacher-training colleges suggests that not more than 50 per cent. of teachers consider religious beliefs to provide the moral basis for their actions.

It has been possible here to mention only the more obvious variations, personal and social, which affect moral growth in a changing society. Such variations complicate the 'normal' process of culture transmission. Some special features of contemporary moral growth which characterize moral maturity in a comparatively unstable society, will now be considered.

V.—THE DEVELOPMENT OF MORAL AUTONOMY.

Democracy, we have seen, encourages a wide degree of deviation from norms of behaviour and belief; its character is to be permissive, to accommodate rather than to reject. Traditional values are, consequently, under continuous challenge in a democracy, except those values which are inherent in the democratic way of life—tolerance, kindliness, justice for example; such values are the abiding values of a democractic culture and can only be challenged by an attack on democracy itself. When to the permissiveness of a democracy are added ease of communication between cultures, and rapid advance in many fields of knowledge, the pattern of culture is sure to become fragmented and diffuse. "Our culture," writes Ruth Benedict, "must deal with cultural standards that go under our eyes and new ones that arise from a shadow upon the horizon."[2]

In such a situation, the process of culture transmission will itself vary from person to person and group to group. The possible permutations of individual inheritance and environmental circumstances are legion ; we may, however, postulate some general groupings which permit the highly complicated picture to be somewhat simplified. Some individuals will be born into a well-organized sub-culture group, or will be attracted into it at a later stage, and will identify themselves with the in-group mores, closing their minds to any criticism of the values they have accepted. These groups are atolls in an ocean of cultural change, comfortable, but constantly threatened by inundation. Other individuals will become victims of the confusion within society and, not having sufficient intelligence or drive to struggle with the confusion, will remain muddled, and devoid of moral clarity. Such individuals compose probably a wide sector of modern society. Some, as has already been noted, will be driven by deprivation of love, esteem or acceptance to adopt anti-social value systems, either egotistically or in groups. In addition, there will be others who acquire a more personal system of values, in part derived from early relationships, in part from subsequent relationships and experiences and, in part, from their own thinking, conscious selection, and re-assessment. This last group is, functionally, the most appropriate to a democratic culture. We have to ask what particular features characterize this kind of moral development over and above the common features of culture transmission.

Little work has so far been undertaken which helps to clarify this point, but it is perhaps legitimate to draw two tentative conclusions. In the first place, moral maturity in a democracy is marked by the attainment of moral autonomy rather than by the complete absorption of any precise code of values. Of course, an act of personal acceptance of values is necessary for all genuine moral development—a development, that is, which is more positive than conformity through fear. Sherif writes : " An effective grasp of the norms of his social groups, of prescribed relations with others around him, comes only after the child has assimilated these norms as *his own*."[19] Yet a child growing up in a changing society needs to go further than this : he has to select between alternatives in values and accept as his standard whichever seems the more worthwhile. Gorer and others have found wide divergencies of belief and behaviour even within in-groups where close conformity might be expected.[20] We may suppose that the attainment of moral autonomy is a gradual process ; early moral acquisition interacts with later experiences to produce the moral outlook at any particular time. It follows that the moral outlook of an adult may be the result of shedding early values without adequate replacements, or the result of rigid adherence to an in-group mores, or the result of transforming early values into a mature moral outlook in which established values are unified with personal reassessments into a viable adult philosophy of life. G. Allport postulates the insights of moral maturity as growing from the rudimentary ethical attitudes of the child in three stages :

" (1) External sanctions give way to internal—a change adequately accounted for by the processes of identification and introjection familiar in Freudian and behavioural theory.

" (2) Experiences of prohibition, fear and ' must ' give way to experiences of preference, self-respect and ' ought.' This shift becomes possible in proportion as the self-image and value systems of the individual develop.

" (3) Specific habits of obedience give way to generic self-guidance, that is to say, to broad schemata of values that confer direction upon conduct."[16]

In the second place, the attainment of moral maturity in a changing society involves searching for a set of trans-cultural values which have universal validity, not absolute values in the metaphysical sense, but aspirations about mankind in terms of which a choice may be made when the culture offers alternatives. Speaking of moral progress, Ginsberg says: " Further progress depends on whether we can formulate a coherent and comprehensive conception of a good common to mankind, whether we can acquire sufficient knowledge of the conditions which are necessary for its realization, and whether we can, in the light of such knowledge, generate a common or co-operative will with sufficient energy to bring these conditions into being."[21]

The relevant point is that those children in our community who are capable of rising above moral apathy, or dependence on cut-and-dried, in-group codes—those capable of moral autonomy—require some sense of involvement with mankind as a whole in order to attain this autonomy.

It is important to notice that the attainment of moral autonomy in the individual does not imply moral anarchy in society. Moral innovation—re-assessment—is a natural feature of a changing society as a whole, and of all but the moribund groups within that society. Moral autonomy permits those persons capable of a creative contribution to the mores of their groups, and of their culture as a whole, to make such a contribution. At present we know a good deal about how conformity is passed on and also about what the significant factors are in the genesis of anti-social behaviour patterns, but regrettably little about the process by which clarity and moral stamina come to be woven together, with tolerance and flexibility of outlook, in the mature personality.

VI.—THE ROLE OF THE SCHOOL.

Although the home is known to be the most powerful formative factor in the moral development of the child, the study of group influences, mainly carried out since the war, has given additional importance to school life in the process of moral growth. In marginal cases, the school's influence is likely to be decisive. During his school years, the child is in daily contact with an organized system of values—the mores of the school—with adult leaders, *in loco parentis*, whom he can, and often does, use as models, and with peer-groups the values of which will be influential whether or not they are consistent with those of the school and home.[22] The process of acquiring values can be affected by the school in four main ways: by reinforcing desirable values already existing in the surrounding community; by reducing the influence of values in the surrounding community which are felt to be undesirable; by promoting the child's self-respect; and by educating his moral insight.

Lippitt and White, among others, have demonstrated the influence upon children's behaviour of classroom relationships. Their experiments showed that twelve-year-old boys changed from restless, aggressive, destructive behaviour to constructive, co-operative behaviour—and back again—in a mere few weeks in response to changes in adult leadership.[23] The inducement of particular kinds of behaviour by particular circumstances does not establish that moral development or regression has occurred, but it does prove the malleability of children of school age. Present views on transfer of training would suggest that, if desirable behaviour is supported by insight into its principles, an enduring impact upon a child's developing 'schemata of values' will be made.[24]

As well as influencing the content of the class-room mores in such a way as to make them socially desirable, a school can also assist by drawing children who are isolated into social relationships with their fellows. The work of Moreno and others suggests that about 10 per cent. of children are isolated at school—denied identification with their peer-group by the rejection of their peers.[25]

Stott and others have shown that solitariness, rejection and isolation appear in many case-histories of delinquency,[9] possibly because moral development becomes fixated at the point where serious isolation is experienced.

Schools offer children clearly-defined goals of striving; they are accordingly in a position to build or undermine the children's self-respect by the way achievement is dispensed among the pupils. Many delinquents have a record of persistent failure at school.[26] Although no research on the influence of school failure upon delinquency, based on scientific comparisons between groups of schools, has so far been completed, we may assume that the success/failure experience of the child at school has repercussions upon his moral development. A child assured of " success in proportion to effort," at school is likely to strive successfully and, so, to maintain his self-respect.

The school can foster moral insight both by drawing children into group discussions on personal and group problems and by offering a content in the curriculum which encourages children to extend their sense of involvement with others. The Eight-Year Study of American Education established that school programmes designed to be broader and more personal than traditional programmes significantly improved the social maturity of students in the immediate post-school years.[27]

In addition to its general influences, the school community can have a particular influence during adolescence when a special opportunity appears to be provided for correcting distortions of personality resulting from earlier years.[28] To quote Kluckhorn again: " In our culture there is often conflict between parental standards and age-group standards."[4] This conflict is liable to be acute during the adolescent years. Children estranged from their parents and, frequently, living away from close contact with other relations, may depend entirely upon their teachers for adult friendliness, help and guidance. Another relevant aspect of the adolescent stage is that stressed by Hollingworth[29] and others—the awakening of altrustic feelings and a yearning to come to terms with the wonder and mystery of existence. This motivation supplies, in the more intelligent at any rate, the energy to weigh problems, and outlooks, to test values already incorporated, to supplement them, and to organize the whole into an incipient philosophy of life.

The influence of Religious Instruction as such on the moral development of children cannot be estimated as little relevant evidence exists. It seems unlikely that values mediated *merely* intellectually will be incorporated into outlook and behaviour. On the other hand, a school which demonstrates the Christian values by its day-to-day way of life and relationships may use Religious Instruction to add insight to experience and will thereby reinforce the influences of the group life. Moreover, seen at its broadest, the total of school influences—intellectual, emotional, social, moral—may be regarded as religious education. Whitehead writes: " A religious education is an education which inculcates duty and reverence. Duty arises from our potential control over the course of events. Where attainable knowledge could have changed the issue, ignorance has the guilt of vice. And the foundation of reverence is this perception, that the present holds within itself the complete sum of existence, backwards and forwards, that whole amplitude of time which is eternity."[30]

The Post-School Period.—The thesis has been put forward that, in a changing democratic society, moral autonomy is the functional mode of moral maturity. This is well illustrated by the experiences young people encounter upon leaving school. As Kinsey has shown for the United States and

Goldberg, Logan[31] and others for this country, sexual behaviour does not conform at all closely to the traditional mores of society : each individual is left to decide for himself how he will behave, with very little social or even group guidance. Jordan and Fisher believe that this need for personal evaluation and decision can assist in the discovery of enduring values. They found a great interest in moral issues among youth club members and a willingness to thrash them out from first principles.[32] It is worth noting also that industrial disputes are fundamentally about moral principles : What is fair ? What is just ? What is true ? Not custom and authority, but argument and negotiation decide. This is the climate in which young people attain adulthood to-day ; moral autonomy is required and assumed.

VII.—The direct influence of psychology.

The insights of psychology into the moral development of the child serve both to describe—albeit incompletely at present—how moral maturity is attained, and also to expose outworn ideas still sufficiently alive within society to affect, through some parents, the moral growth of children, and to influence directly the developing moral outlook of some children. The contribution of psychology in this particular sphere is constantly growing. To give three examples :

(1) It was formerly assumed that the moral powers of the child were inherent—a conscience was innate along with original sin. Consequently, moral development was treated as unlike all other forms of development in that early stumblings were considered to be a proof of evil rather than a mark of immaturity. Psychological research has shown that moral skill is comparable to any other skill, and that the advance from infantile egocentricity to adult maturity should be a continuous process which can be expected to follow a natural course if properly nourished by the right relationships and experiences.

(2) Prior to psychological inquiry, virtue was held to be the outcome of repressing the bad—conditioning it out of existence. We now realise that the sources of virtue and vice are the same—the basic energies of man. Moral development is best served not by splitting these energies into aspects of the self accepted as ' me ' and others rejected as ' not me ' but by using and socializing all the energies of the psyche in the attainment of an integrated personality.

(3) A similar social confusion failed to notice the difference between infantile egocentricity and ' the self.' Moralists urged us to break the will of children and to seek always to be someone other than ourselves. We now see that the uniqueness of each person is what is most valuable about him—that self-development is the true educational aim. Thus, Hadfield concludes that moral maturity requires us to know ourselves, accept ourselves and be ourselves.[33]

Such insights of psychology—many more could be cited—are relevant to the subject in hand because, while bringing clarity to a muddled area of human thought, they inevitably further confuse the climate of values surrounding young people. The in-group mores of our culture are under test from the influx of other mores and new knowledge in general, and also from the impact of psychological inquiry into the nature of man. The more intelligent young people of our community appear to become interested in psychology between the ages of sixteen and twenty. To the extent that this interest leads to a reassessment of moral ideas still active in the community, it is a formative influence in the moral development of this sector of the community.

VIII.—CONCLUSION.

The purpose of this paper has been to set out some lines of evidence on the way values are transmitted from generation to generation and to relate this evidence to the moral development of children in a rapidly-changing society. A number of important factors in moral development have had to be omitted, or almost omitted, in offering this over-simplified description of the process of moral development. For example, no specific mention has been made of guilt, and anxiety about guilt, as an enduring influence in character; the effects upon young people of films, radio, television, and the popular press have not been discussed, nor the part that symbols play in organizing and sustaining schemata of values, nor the influence upon society of the decline in ritual and ceremony, nor the role of the ego-ideal in sustaining effort. All these factors are important but their inclusion would have made the paper excessively long.

The paper suggests that the process of moral development in a changing society may usefully be considered in terms of the modifications imposed on the uncomplicated process of value transmission by the widespread variations which are found in western culture to-day. It further suggests that a degree of moral autonomy—a capacity personally to reassess values which are encountered in experience—is a necessary factor in moral maturity in a changing society.

The recent studies of group psychology give additional significance to the role of the school in culture transmission.

The task of psychologists in relation to moral development would appear to be a dual one: to study the process of moral development as precisely as possible, and to clarify confused ideas in the light of current psychological knowledge about human growth and behaviour.

I.—REFERENCES.

[1] READ, M. (1955). *Education and Social Change in Tropical Areas.* Nelson.
[2] BENEDICT, R. (1935). *Patterns of Culture.* Kegan Paul.
[3] MALINOWSKI, B. (1926). *Crime and Custom in Savage Society.* Kegan Paul.
[4] KLUCKHORN, C. (1950). *Mirror for Man.* Harrap.
[5] HORNEY, K. (1951). *Neurosis and Human Growth.* Kegan Paul.
[6] FLUGEL, J. C. (1945). *Man, Morals and Society.* Duckworth.
[7] ISAACS, S. (1929). *The Nursery Years.* Routledge.
[8] ADLER, A. (1938). *Social Interest.* Faber and Faber.
[9] STOTT, D. H. (1950). *Delinquency and Human Nature.* Carnegie United Kingdom Trust.
[10] MARTIN, P. W. (1955). *An Experiment in Depth.* Kegan Paul.
[11] FROMM, E. (1949). *Man for Himself.* Kegan Paul.
[12] BURN, M. (1956). *Mr. Lyward's Answer.* Hamish Hamilton.
[13] COHEN, A. K. (1956). *Delinquent Boys—The Culture of the Gang.* Kegan Paul.
[14] PIAGET, J. (1932) *The Moral Judgment of the Child.* Kegan Paul.
[15] GESELL, A., ILG, F. L., and AMES, L. B. (1956). *Youth.* Hamish Hamilton.
[16] ALLPORT, G. W. (1955). *Becoming.* Yale University Press.
[17] BOWLBY, J. (1947). *Forty-Four Juvenile Thieves.* Baillière, Tindall and Cox.
BOWLBY, J. (1951). *Maternal Care and Mental Health.* World Health Organization.
[18] WALL, W. D. (1948). *The Adolescent Child.* Methuen.
[19] SHERIF, M. (1948). *An Outline of Social Psychology.* Harper.
[20] GORER, G. (1955). *Exploring English Character.* The Cresset Press.
[21] GINSBERG, M. (1944). *Moral Progress.* (Frazer Lecture, 1944). Glasgow: Jackson.
[22] CRUZE, W. W. (1953). *Adolescent Psychology and Development.* New York: Ronald Press.
[23] LIPPITT, R., and WHITE, R. K. (1947). An experimental study of leadership and group life. *Readings in Social Psychology* (edit. Newcomb and Hartley). New York: Holt.
[24] BARTLETT, F. C. (1951). *The Mind at Work and Play.* Allen and Unwin.
[25] MORENO, J. L. (1934). *Who Shall Survive?* Beacon House.
[26] BURT, C. (1925). *The Young Delinquent.* University of London Press.
[27] *Adventures in American Education* (1942). Harper.

[28] HURLOCK, E. B. (1949). *Adolescent Development.* New York: McGraw-Hill.
[29] HOLLINGWORTH, L. S. (1930). *The Psychology of the Adolescent.* Staples.
[30] WHITEHEAD, N. W. (1932). *The Aims of Education.* Williams and Norgate.
[31] LOGAN, R. F. L., and GOLDBERG, E. M. (1953). Rising eighteen in a London suburb. *Brit. J. Sociol.*, 4, 323-345.
[32] JORDAN, G. W., and FIDHER, E. M. (1955). *Self-Portrait of Youth.* Heinemann.
[33] HADFIELD, J. A. (1923). *Psychology and Morals.* Methuen.

(*Manuscript received September* 11*th*, 1956)

THE OBJECTIVE USE OF A PROJECTIVE TECHNIQUE, ILLUSTRATED BY A STUDY OF THE DIFFERENCES IN ATTITUDES BETWEEN PUPILS OF GRAMMAR SCHOOLS AND OF SECONDARY MODERN SCHOOLS

By EVA BENE

(Institute of Psychiatry, Maudsley Hospital, London,
formerly London School of Economics).

SUMMARY.—This paper is concerned with a new method for coding responses to the sentence completion test which makes it possible to use this projective technique for objective investigations of the attitudes of large samples.

1.—A description of the sentence completion test is given with illustrations of, and a discussion of the assumptions underlying, its use.

2.—The proposed coding method is described; each response is coded by two symbols, one representing the attitude expressed in it, and another standing for the object towards which the attitude has been expressed. After the responses have been coded in this manner, they can be treated statistically without much loss or distortion of their original meaning.

3.—The applicability of the coding method is demonstrated by an investigation into differences in attitudes between pupils of grammar and of secondary modern schools. Significant differences were obtained between the school groups, which indicated that grammar school boys are more ambitious and have less positive and more negative attitudes towards their environment than have secondary modern school boys. Some of the possible reasons for these differences were examined and excluded, others remain to be investigated.

4.—Problems connected with the prediction of manifest behaviour from test responses are discussed.

I.—INTRODUCTION.

INVESTIGATIONS of the emotional attitudes of large samples are handicapped by the scarcity of suitable techniques. Questionnaires and check-lists give results in an objective manner but prevent the subjects from giving spontaneous expressions to their feelings. Projective tests allow the subjects freedom of expression but are difficult to use in an objective way. The interpretation of the records takes much time and skill and results in clinical pictures which are often too complex and varied for orderly classification and comparison.

The purpose of this paper is to show how the sentence completion test can be used in a manner in which the advantages ascribed to projective techniques are combined with those of more objective methods.

II.—DESCRIPTION OF THE SENTENCE COMPLETION TEST.

The sentence completion test consists of incomplete sentences which the subjects are asked to finish. The words or short phrases which form the items are chosen to elicit responses which may reveal the subjects' attitudes, and especially those which are relevant to the particular investigation. The word 'attitude' is to be understood in this context in its broadest sense including what are variously called affects, dispositions, sentiments and feelings. Discussions of the history of the test and its various forms can be found in publications by Bell (1948), by Anderson and Anderson (1951) and by Brower and Abt (1952).

The value of the sentence completion test as a diagnostic tool has been shown in a number of studies. Rohde (1946) compared sentence completion scores with personality ratings based on interviews and school records and obtained correlations ranging around 0·80. Rotter and Willerman (1947) and Rotter, Rafferty and Schachtitz (1949) used clinical assessments as criteria and obtained lower, although still significant, correlations. Rotter, Rafferty and Lotsof (1954) found that the test discriminated effectively between maladjusted children referred to clinics and presumably normal school children. Levy (1952) was interested in finding out what tests are considered to be most useful in clinical work, and found that his sample of forty-two clinical psychologists seemed to think that after the Rorschach and the TAT the sentence completion test is their most useful instrument.

The value of the sentence completion test as a research tool has also been demonstrated. Kelly and Fiske (1951), for example, state that it is a promising tool in the prediction of the vocational aptitude of clinical psychologists. Kimball (1952) showed that the sentence completion technique can be successfully used to test hypotheses derived from clinical studies. Lindgren (1954) investigated the influence of learning experience on the attitudes of students and suggested that " . . . researchers may be able to chart the ebb and flow of attitudinal changes in groups of various sorts."

A comparison made between sentence completion records and personality assessments based on overt behaviour gave less satisfactory results than did the studies mentioned above. Symonds (1947) compared soldiers' sentence completion records with their army reports and found almost as much disagreement as agreement. He concluded that the army reports dealt with present personality, behaviour and attitudes, whereas sentence completions showed underlying trends, phantasies and unconscious impulses, which may or may not have counterparts in conscious attitudes and behaviour.

Ever since the sentence completion test has been used, the assumption has been made that the record of a subject may contain not only conscious expressions of his attitudes but also indications of his covert feelings. The subject may reveal attitudes of whose existence he knows but yet be unaware that he is expressing them in the test situation; or he may be quite unaware of the existence of the attitudes which are manifested in his responses. Thus, the subject's responses may be expressions of his public feelings, private feelings, or unconscious feelings. Often, some of the responses contained in a record are found to be contradictory, and it is held that such contradictions probably reflect conflicts in the emotional life of the subject. These assumptions are commonly made for most projective techniques and have been discussed in an extensive literature, including such recent publications as those of Abt and Bellak (1950), Anderson and Anderson (1951), and Brower and Abt (1952). If these assumptions are accepted, the sentence completion test can be regarded as a useful technique for obtaining samples of attitudes.

III.—THE CODING METHOD.

Compared with most other projective techniques, the sentence completion test has the advantage that it is as effective as a group test as it is as an individual test, and that it can be administered easily to large samples. The test material is simple, the instructions given to the subjects are easy to follow; it can usually be given as a paper-and-pencil test, and if the subjects cannot write, they can dictate their responses. The major difficulties it presents concern the meaningful organisation of the data.

In individual diagnostic work the records are usually interpreted in a global manner and the quality of the interpretation depends on the clinical understanding of the interpreter. Holsopple and Miale (1954) suggest a sentence by sentence interpretation from which they say the outstanding properties of the record will eventually emerge. Shor (1946) analyses the responses in terms of areas of resistance, evasion, rejection, recurrent themes, typical associations and specifity of attitudes. Some investigators have proposed quantitative scoring methods ; Rohde (1946) classifies the responses in terms of thirty-nine personality categories based on Murray's system of needs, presses and inner states, and quantifies the data by counting the frequency and estimating, on a three-point scale, the intensity with which each variable was expressed by the subject. Rotter and Willerman (1947) suggest rating the responses as healthy, neutral or unhealthy on a scale ranging from plus three to minus three and give scoring standards for each of their forty items. Forer (1950) has check lists of attitude areas for responses to various groups of items, and Lindgren (1954) scores the responses in terms of a three-point scale of acceptance and rejection.

In the sentence completion test the subject may give any response to any item and it is difficult to know how to score a response if it only fits the scoring categories assigned to a different item and not those proper to itself. Therefore, the coding method suggested in this paper uses categories which apply to responses made to all items in the test. Other advantages of the present coding system are that the investigator is not tied to a specific approach to personality which might or might not be acceptable to him ; no difficult judgments have to be made such as evaluations of responses along a numerical scale ; and the responses do not lose most of their content by being rated along a single continuum or classified into a limited number of categories.

Whereas sentence completion responses are usually coded as units, the method suggested in this paper involves two sets of codes ; one for the objects and one for the attitudes that occur in the responses. To have two sets of codes has the advantage that responses can be recorded with only little loss or distortion of their original meaning, and that coding categories can be devised which can apply to responses made to any items in the test. The objects of the responses are classified into such categories as the self, various other people, institutions and activities ; the attitudes into such categories as positive and negative feelings, anxiety, guilt, aspirations. Thus, each response is represented by a pair of symbols, one standing for the object and one for the attitude towards that object. In addition, there are symbols for responses which express no attitudes, for responses that contain attitudes or objects that occurred too rarely to be included in the system of classification and for items which are left blank by the subjects. When all the responses have been coded by the proper combination of symbols, one can establish how many subjects have expressed a given attitude towards a particular object by counting the number of records in which the appropriate pair of symbols occurs ; one can also count the frequency with which the various pairs of symbols are contained in each record.

Sentence completion responses coded according to this method can be examined quantitatively in several ways. One can examine the relative frequency with which various emotional attitudes were expressed by a sample. One can compare the emotional attitudes of two samples by comparing the frequency with which the various pairs of codes occur in both sets of records. It is also possible to compare the attitudes of one individual with those of a sample by observing how frequently his codes occur in the records of the sample.

IV.—THE OBJECTIVE USE OF THE SENTENCE COMPLETION TEST ILLUSTRATED BY A STUDY OF THE DIFFERENCES IN ATTITUDES BETWEEN PUPILS OF GRAMMAR SCHOOLS AND OF SECONDARY MODERN SCHOOLS.

This section shows how the coding technique was employed to make objective comparisons between the emotional attitudes of various samples of adolescent boys.

After comparing the environment in which the average grammar school boy spends his life with that of the average boy who attends a secondary modern school, many would judge the environment of the grammar school boy to be superior. The parents of grammar school boys are on the whole better educated and more intelligent, and they belong to a higher social class ; grammar schools carry more prestige and give both better education and better opportunities for future success than do secondary modern schools. However, actual comparison between the emotional attitudes expressed in sentence completion responses of grammar school and of modern school boys indicates that it is the grammar school boys who feel the more resentment towards their environment.

This statement is based on data collected by the Social Psychology Department of the London School of Economics for the purposes of an investigation into various attitudes of adolescent boys. The subjects were between the ages of thirteen and fifteen, they were in the third year of their secondary education, and were drawn from four grammar and four secondary modern schools in the Greater London area. One of the grammar schools was located in a predominantly middle class district, another in a predominantly working class district, and the other two in districts in which the social classes were about equally represented. The general surroundings of the secondary modern schools were similar to those of the grammar schools. The distribution of the subjects between the two types of schools is shown in Table 1.

TABLE 1

SCHOOL DISTRIBUTION OF SUBJECTS.

	N
Grammar School Sample	317
Secondary Modern School Sample	301
Total	618

The sentence completion test which was used for the investigation of emotional attitudes was made up of items which, it was hoped, would enable young adolescent boys to express their feelings about themselves and about significant people, institutions, and activities in their lives. Some of these items had been used before by Tendler (1930), Rohde and Hildreth (1940), Shor (1946) and others. The test was so constructed that to some items the boys had to express attitudes towards given objects, to others they had to find objects to give attitudes, and to a third type they had to supply both the objects and the attitudes. The items are given in the following list :

LIST OF ITEMS

Items which determined the objects :

Father . . .
Mother . . .
If only my parents . . .
My parents think that I . . .
Brothers and sisters . . .
When I am at home . . .

Grownups . . .
The world . . .
Being my age is . . .
Other children . . .
Boys are . . .
Others boys probably say about me that . . .
Girls . . .
Marriage . . .
The teachers . . .
School is . . .
My schoolwork . . .
Sports . . .

Items which suggested attitudes :
I am glad that I . . .
I wish . . .
I hope . . .
I like it when . . .
I am happy when . . .
I hate . . .
I cannot . . .
I am sorry when . . .
If only I had not . . .
If only I were not so . . .
What annoys me is . . .

Items with no specific content :
I secretly . . .
There are times when . . .
Sometimes I think that . . .
I wonder if I shall . . .
When I am alone I . . .
When I am twenty-five I . . .

In the actual test the three kinds of items were given in a mixed order. The boys were instructed to finish the sentences as quickly as possible and not to worry about style or spelling. The test was administered in classrooms but no teacher or other school authority was present and the boys did their work anonymously.

The coding categories were devised on a theoretical basis, but it was found that they needed some modifications to fit the objects and attitudes that actually occurred in the records. Some categories had to be pooled because objects or attitudes as expressed by the boys were too ambiguous to fit clearly into one or the other ; some remained empty and were omitted ; and a few new categories had to be made up for objects and attitudes that were found unexpectedly in the responses. To facilitate the use of Hollerith cards, digits from 0 to 9 and the letters x and y were used for coding symbols. The first digit in each pair stood for the object and the second for the attitude of a given response. Although the attitudes were classified into more than twelve categories, they were represented by only twelve symbols and the meaning of the second symbol in each pair was specified by the meaning of the first. In this manner the responses could be punched by the use of two columns only and space was saved on the Hollerith cards. The coding categories and their symbols are presented in Table 2. Examples of the coding responses are given in Table 3.

TABLE 2

CODING CATEGORIES AND THEIR SYMBOLS.

A.—*Object of Response, represented by First Symbol in each Pair.*

0 = parents, home
1 = siblings
2 = authority figures, such as teachers
3 = boys, children, people
4 = girls, marriage
5 = the self
6 = school
7 = sports, organised games
8 = miscellaneous

From the thirty-eight coding categories dealing with attitudes in relation to the environment, fourteen showed significant differences between the school groups. These are given in Tables 5 and 6. The Tables show that a smaller percentage of grammar school boys expressed positive, and a larger percentage negative, attitudes towards the environment than was the case with the modern school boys. The fact that all the significant differences that occurred are consistent in their directions is an added indication that they did not occur by chance.

It is interesting to note that no significant differences were found between the school groups in the boys' own attitudes towards parents and teachers but only in the reported attitudes of parents and teachers towards themselves. These differences may be due to the fact that the parents and teachers of grammar school boys are more authoritarian. One could also speculate that some of the grammar school boys used projection as a defence against the guilt they would experience if they admitted having negative feelings, or lacking positive feelings, towards these most important adults. In the case of siblings and girls, who can be disliked without much guilt, significant differences occurred on those coding categories that had to do with the boys' own feelings.

The coding categories which showed differences in ambition are given in Table 7. There were altogether six categories for responses expressing ambition, and significant differences were found on four. The direction of all four differences points towards greater ambition on the part of the grammar school boys and seems to show that these boys are under greater pressure by their parents and have a higher level of aspiration than is the case with boys who attend secondary modern schools.

TABLE 5

SIGNIFICANT DIFFERENCES IN POSITIVE EMOTIONAL ATTITUDES TO THE ENVIRONMENT BETWEEN THE GRAMMAR SCHOOL AND THE MODERN SCHOOL SAMPLES.

	Grammar	Modern	Level of sign
Positive feelings felt toward :	%	%	
parents, home	84	84	—
authority figures such as teachers	41	47	—
siblings	28	38	0·01
boys, children, people	44	47	—
girls	81	89	0·01
school	72	81	0·01
sports, games	75	83	0·05
the world as a whole	9	13	—
Positive feelings felt as coming from :			
parents, home	66	74	0·05
authority figures such as teachers	11	21	0·003
boys, children, people	8	8	—
Desire felt to help or please :			
parents, at home	34	37	—
boys, children, people	9	6	—

TABLE 6

SIGNIFICANT DIFFERENCES IN NEGATIVE EMOTIONAL ATTITUDES TO THE ENVIRONMENT BETWEEN THE GRAMMAR SCHOOL AND THE MODERN SCHOOL SAMPLES.

	Grammar	Modern	Level of sign
Negative feelings felt toward :	%	%	
parents, home	21	14	—
authorty figures such as teachers	33	30	—
siblings	47	30	0·003
boys, children, people	22	27	—
girls	61	52	0·05
school	62	48	0·003
sports, games	13	7	—
the world as a whole	48	27	0·003
Negative feelings felt as coming from :			
parents, home	50	53	—
authority figures such as teachers	50	36	0·003
siblings	3	5	—
boys, children, people	17	11	—
Strong hatred, aggression felt toward, complete rejection of :			
parents, home	6	8	—
authority figures such as teachers	8	6	—
siblings	3	5	—
boys, children, people	8	11	—
Resentment due to aggressive behaviour of :			
parents	9	10	—
authority figures such as teachers	13	13	—
boys, children, people	19	20	—
Feelings of anxiety, guilt, due to :			
parents	12	9	—
boys, children, people	13	3	0·003
school	27	12	0·003
Criticism of manners, behaviour of :			
parents	4	5	—
authority figures such as teachers	12	12	—
boys, children, people	41	30	0·01

TABLE 7

SIGNIFICANT DIFFERENCES IN AMBITION BETWEEN THE GRAMMAR SCHOOL AND MODERN SCHOOL SAMPLES.

	Grammar	Modern	Level of sign
	%	%	
Wants to do well at school	50	32	0·003
Wants to do well at sports, games	16	18	—
Is dissatisfied with his school achievements	53	40	0·003
Wants to become rich, well off	31	18	0·003
Feels under pressure from his parents	53	37	0·003
Feels under pressure from his teachers	5	2	—

The finding that boys who attend grammar schools have higher levels of aspiration than have boys who attend secondary modern schools, is not very surprising ; however, that they have less positive and more negative attitudes seems to require some explanation. The differences in attitudes between the school groups might be due some of the following reasons :

(1) Differences in school background. Grammar schools, compared with secondary modern schools, make greater demands on their pupils ; their pupils are under greater pressure, feel more frustrated and, therefore, have fewer positive attitudes.

(2) Differences in intelligence. Pupils of grammar schools are more intelligent and, therefore, more critical than are pupils of modern schools.

(3) Differences in level of aspiration. Children whose level of aspiration is high and who are expected by their parents to do well in school will, other factors being equal, more frequently find their way into grammar school than will other children. These children will strain themselves to achieve success, will be anxious lest they fail and, therefore, see the world in less pleasant terms than will other children.

(4) Differences in social class. Grammar schools contain a larger proportion of middle class children than do secondary modern schools and it is possible that membership in the middle class makes for less positive attitudes.

(5) Differences in family size. Children who attend secondary modern schools more frequently come from large families and, therefore, have better opportunities to learn to get along with people.

These possible explanations can all be expressed in terms of hypotheses that can be tested by the use of the sentence completion test. One could, for instance, give the test to the same children twice ; first, while they attend primary schools, and again after they have attended their various secondary schools for some time. If the results of the second test show differences between the attitudes of grammar school and modern school children which cannot be found when they were first tested, it is likely that the differences between their attitudes are due to differences in their school background. Similarly, one could also test the hypotheses that higher intelligence or higher level of aspiration is related to more negative and less positive attitudes to the environment.

In fact, the hypotheses that the differences in attitudes are due to differences in social class or family size have been tested. Davis and Havighurst (1947) suggest that middle class children have a stricter upbringing with more rules and prohibitions and less sensual gratifications than have working class children. These differences in early childhood frustrations and gratifications may account for the more negative attitudes of the grammar school sample, as it contained a much higher proportion of middle class boys than did the modern school sample.* When, however, the middle and the working class boys were compared with one another within each of the two sorts of school, little more than chance differences were found between their attitudes.

Another of the possible explanations is that modern school boys have more opportunities to learn to get along with people, because on the whole, they have more siblings. However, the results of comparisons made between boys from families with two children or less and boys from families with three children or more did not indicate that the differences found between the school groups were due to differences in family size. What they are due to seems to be an important enough question to warrant further research.

---

* The social class distribution of the sample was carried out on the basis of a modified version of the Hall Caradog-Jones scale (1950), of occupational prestige, and has been described by Himmelweit, Halsey and Oppenheim (1952).

V.—CONCLUSIONS.

An example has been given of the proposed coding method in which coding categories applied to responses of adolescent boys have been demonstrated. However, the method does not involve a fixed number of categories with fixed contents, but is rather a scheme, on the basis of which each investigator can develop that system of classification of responses which is most suitable for his work. It is the opinion of the writer that there is no one set of coding categories that would suit all investigations in which the sentence completion test can be used, since the content of the responses varies with the age, sex, and other characteristics of the sample and also depends on the items employed. Once the categories have been decided upon the responses can be coded with relative speed and ease, and the bulk of the statistical work can be done by Hollerith machines.

In the investigation described in this paper, the significant differences that were obtained fell into a consistent, meaningful pattern ; this seems to show that they are not artificats of the test situation but originate in real differences between the boys' attitudes. The results indicate that grammar school boys have less positive and more negative attitudes than have modern school boys ; the question arises whether they also indicate that grammar school boys are less friendly and more aggressive in their overt behaviour. It seems important to point out that sentence completion data give no information by which one could tell whether and to what extent differences in attitudes are offset by differences in other personality and environmental factors. Many more grammar school than modern school boys have a middle class upbringing, better manners, more self-discipline, and less conflict of values between their homes and schools. They are more intelligent, have greater expectations for their future, and in addition, their environment is of a different kind and presents different possibilites for social interaction. We have seen that the sentence completion test, like other psychological techniques, can furnish valuable information about some aspects of an individual's personality ; inferences from these aspects to his total personality would have to be made on the basis of a theory of personality which, as Wyatt (1952) has pointed out, lies above and beyond the test itself. However, even if we knew all about an individual's personality we would not be able to predict how he is going to act under a specific set of concrete circumstances, but only what his tendencies towards behaviour are. To predict actual behaviour requires a knowledge of the total external situation as well, as behaviour depends on a complex interaction between the total personality and its total environment. On the basis of sentence completion results alone one can tell very little about overt behaviour. If, for instance, a boy's responses express hostility towards his father, we can assume that his hostility will influence his behaviour in some way ; but unless we have much more additional information about him we can not tell whether he tends to express his aggression directly through hostile conduct towards his father, tends to displace it towards some other person, or turn it against himself.

Kelly (1954) has asked why psychologists use techniques for which there is little or no evidence of predictive validity, similarly, one might question what useful research purposes the sentence completion test may serve considering its limitations in predicting manifest behaviour. The writer believes that the test is a useful tool for investigations concerned with the influence various personality and social factors may have on people's attitudes ; what can be tested objectively is the existence of constant relationships between attitudes and such personality variables as age, sex, intelligence, state of health, and between attitudes and environmental variables such as various kinds of home, school, social class and ethnic backgrounds. The contention of this paper is that investigations of such relationships are important both for social and for clinical psychology and that the suggested coding method can help investigators to use the rich samples of attitudes the sentence completion test can provide.

* * * * * *

ACKNOWLEDGEMENTS.—The author wishes to acknowledge her indebtedness to Dr. Hilde Himmelweit for her kind permission to use the data required for this study and to Mr. Alan Tyson for his many helpful suggestions. Her thanks are also due to the Division of Social and Demographic Research of the London School of Economics for financial support for the study and to the head masters and pupils of the schools involved for their generous co-operation.

VI.—REFERENCES.

ABT, L. E., and BELLAK, L. (1950). *Projective Psychology*. New York: Alfred A. Knopf.
ANDERSON, H. H., and ANDERSON, G. L. (1951). *An Introduction to Projective Techniques* New York: Prentice-Hall, Inc.
BELL, J. E. (1948). *Projective Techniques*. New York: Longmans, Green and Co.
BROWER, D., and ABT, L. E. (1952). *Progress in Clinical Psychology*. New York: Grune and Stratton.
DAVIS, A., and DOLLARD, J. (1940). *Children of Bondage*. Washington, D.C.: American Council on Education.
DAVIS, A. (1947). Socialisation and adolescent personality. (In Newcomb, Hartley, *et al.*, *Readings in Social Psychology*. New York: Henry Holt and Co.).
DAVIS, A. ,and HAVIGHURST, R. J. (1947). *Father of Man*. Boston: Houghton Mifflin Co.
FORER, B R. (1950). A structured sentence completion test. *J. Proj. Tech.*, 14, 15-30.
HALL, J., and CARADOG-JONES, D. (1950). The social grading of occupations. *Brit. J. Sociol*, 1, 31-55.
HIMMELWEIT, H. T., HALSEY, A. H., and OPPENHEIM, A. N. (1952). The views of adolescents on some aspects of the social class structure. *Brit. J. Sociol.*, 3, 148-172.
HOLSOPPLE, J. Q., and MIALE, F. R. (1954). *Sentence Completion*. Springfield, Ill.: C. C. Thomas.
KELLY, E. L., and FISKE, D. W. (1951). *The Prediction of Performance in Clinical Psychology*. Ann Arbor, Mich.: University of Michigan Press.
KELLY, E. L. (1954). Theory and techniques of assessment. *Ann. Rev. Psychol.*, 5, 281-307.
KIMBALL, B. (1952). The sentence completion technique in a study of scholastic underachievement. *J. Consult. Psychol.*, 16, 353-358.
LEVY, S. (1952). Sentence completion and word association test. (In Brower and Abt, ed., *Progress in Clinical Psychology*. New York: Grune and Stratton.)
LINDGREN, H. C. (1954). The use of a sentence completion test in measuring attitudinal changes among college freshmen. *J. Soc. Psychol.*, 40, 79-92.
ROHDE, A. R., and HILDRETH, C. (1940). *Sentence Completions*. New York: The Psychological Corporation.
ROHDE, A. R. (1946). Explorations in personality by the sentence completion method. *J. Appl. Psychol.*, 30, 169-181.
ROTTER, J. B., and WILLERMAN, B. (1947). The incomplete sentences test as a method of studying personality. *J. Consult. Psychol.*, 11, 43-48.
ROTTER, J. B., RAFFERTY, J. E., and SCHACHTITZ, E. (1949). Validation of the Rotter incomplete sentences blank for college screening. *J. Consult. Psychol.*, 13, 348-356.
ROTTER, J. B., RAFFERTY, J. E., and LOTSOF, A. B. (1954). The validity of the Rotter incomplete sentences blank, high school form. *J. Consult. Psychol.*, 18, 105-111.
SHOR, J. (1946). Report on a verbal projective technique. *J. Clin. Psychol.*, 2, 279-282.
SYMONDS, P. M. (1947). The sentence completion test as a projective technique. *J. Abnorm. and Soc. Psychol.*, 42, 320-329.
TENDLER, A. D. (1930). A preliminary report on a test for emotional insight. *J. Appl. Psychol.*, 14, 123-136.
WYATT, F. (1952). Prediction in the Rorschach Test. *J. Proj. Tech.*, 16, 252-258.
ZUBIN, J. (1939). Nomographs for determining the significance of the differences between frequencies of events in two contrasted series of groups. *Am. Stat. Assoc.*, 34, 539-544.

(*Manuscript received, June 10th, 1956*).

ON THE RELATION BETWEEN THE SCHOOL JOURNEY AND SOCIAL AND EMOTIONAL ADJUSTMENT IN RURAL INFANT CHILDREN*

By TERENCE LEE
(*Department of Psychology, University of St. Andrews.*)

SUMMARY. 1.—883 children from 57 rural primary schools were assessed by their teachers on ten dimensions of personality.

2.—Comparisons between children with (*a*) short walking, (*b*) long walking, (*c*) short transport, and (*d*) long transport journeys to school showed differences for individual traits and for a combined index, suggesting a relationship between long journeys and maladjustment.

3.—For approximately equal times, transport journeys are associated with poorer adjustment than walking journeys.

4.—Boys in the sample show poorer adjustment than girls and greater differences for length of journey.

5.—The apparent effect lessens with age in respect of walking but not of transport journeys.

6.—The evidence is discussed in relation to explanatory hypotheses of 'family difference,' fatigue and maternal separation. It is concluded that the latter is most in accord with the data.

7.—In particular, it is concluded that the effect of school journeys should be studied through the child's perception of his mother's accessibility, a variable linking maternal separation with maternal deprivation.

I.—INTRODUCTION.

THE task of improving and re-organising education in rural areas is be-devilled by a sparse and sometimes falling population. One aid to solution which has been widely adopted, though often disputed, is to close small primary schools and to transport their children to neighbouring villages.

The many pros and cons of this policy have been debated during the past ten years in the arenas both of education and of politics and cannot be re-appraised here. It must be enough to mention four main arguments.

The first is the least equivocal—the economic burden of maintaining and improving, if it is avoidable, a multiplicity of small units. Secondly, the educational virtues of a wider curriculum and varied staff are contrasted with those of local, more individual attention. Thirdly, arguments range upon the debilitating effects which the loss of a school may have upon village social life; and finally, on whether a longer journey to school may not have some direct harmful effect upon the child itself.

The extent of closure may be assessed from figures for the County of Devon. Since 1900, approximately 170 schools (30 per cent. of the total) have been closed, while the population has increased by 20 per cent. These figures are surprising, but they refer, of course, to an area where the problem of small dispersed units is particularly serious.

School closure is a complex social issue, which requires to be broken down into component problems that are amenable to research, with a cautious awareness that evidence for any one of these does not resolve the total issue.

---

* The present report is a condensed version of the (unpublished) Research Bulletin No. 1 of the Institute of Education, University of Exeter, where the work was carried out.

The particular problem to be considered in this paper is the possibility that closure, through its inevitable increase in the distance that children must travel, has a deleterious effect upon early social and emotional development.

In homely terms, it is concerned with the argument between those people who " will not tolerate tiny tots beating about in all winds and weathers over miles of country roads " (" The Farmers' Letter," *Western Gazette*) ; and those who reply that in their young day children walked miles without " taking harm," and now they ride in comfort at the ratepayer's expense.

II.—OUTLINE OF THE INVESTIGATION.

Primary school teachers in Devon were asked to assess their children on a number of personality dimensions in order that the mean scale positions of groups of children with differing forms and lengths of school journey could be compared.

An index of social and emotional adjustment was computed by combining, for each child, his scale position on these dimensions. This is used to illustrate the trend of relationship with increasing distances of walking journey, and with increasing lengths of time spent in bus and other transport to school. Differences between these two main forms of journey with age and sex are also shown.

The evidence, which suggests that long school journeys are associated with maladjustment* is discussed in relation to three explanatory hypotheses: (i) genetic or child-rearing differences between village and ' remote ' families (i.e., the relationship is concomitant, not causal) ; (ii) physical fatigue ; (iii) maternal separation.

III.—METHOD.

To visit an adequate number of small primary schools in Devon, spending some time at each, would have been costly and time consuming. It would also have been difficult to conceal the purpose of the investigation.

A solution was found by providing coaches which followed routes round the countryside collecting teachers in much the same way as they usually collect children. Teachers were brought to meetings, of which two were held, the first in Exeter and the second in South Molton. An excellent response was gained in this way.

The invitation to teachers explained that they would be asked to provide, at the meeting, certain information about the backgrounds and ' classroom behaviour' of their children "for an investigation of one of the possible causes of maladjustment." Nothing further was said about the particular objective of the work until after the teachers had recorded their judgments.

The Area.—A large, thinly populated, agricultural area was selected, comprising about 650 square miles of mid-Devon, about one-quarter of the total area of the county. It extends northwards from the fringes of Dartmoor to a line drawn from Barnstaple on the North-West coast, to North Molton, near the Somerset border ; westwards to Halwill and Lewtrenchard, near the Cornwall border, and in a South-Easterly direction, to within a few miles of Exeter.

It would have been unsatisfactory if the area sampled had been influenced by a dependence on main road routes. In the case of each meeting, therefore, two coaches started in opposite directions from a single point on the far side of a circle—and teachers from inside and outside this circle made their way by bus, private car or taxi to stopping points on the main routes which converged on the meeting place. By this means, the whole area was covered and where schools were not represented, it was for reasons other than their remoteness from communications.

---

* It is pointed out that the term ' maladjustment ' is not used here in the official sense, denoting one of the categories of handicapped children, but in the sense in which children can be assessed as more or less well adjusted on a continuous scale, with reference to certain dimensions of behaviour. All the comparisons are relative and the methods used do not permit of absolute judgments.

The number of primary schools in the area so defined is sixty-six. Of these, the teachers of three were reluctant to attend, four were prevented by sickness, one head teacher had recently died and one was a ' temporary supply.' Teachers from a total of fifty-seven schools provided data for the investigation.

The Sample of Children.—The sample consisted basically of all six and seven year-old children from the schools in the area, with some slight variations of which details are given in Lee (1956). At the second meeting only, in order to gain some data for eight-year-old children, some teachers from small schools were asked to include these, providing a small sub-sample.

Assessments were made of a total of 883 children, 467 boys and 416 girls ; 409 children were aged six, 428 aged seven, and 46 aged eight years old.

The age range was restricted on the expectation that if an effect was to be found at all, it would be most apparent among young children. On the other hand, five-year-old children were excluded in order to minimise any confusion with the effects of the initial home/school transition.

Procedure at the Research Meetings.—Approximately fifty teachers attended each meeting and were seated at tables with the rating apparatus which is described below. From schools with a staff of more than one, many head teachers accepted the invitation and accompanied their assistants who had the daily care of the children and who were primarily responsible for making the assessments.*

After a brief introduction, instructions were given, including the usual warnings about the dangers inherent in the rating process. No attempt was made to mould the distributions except that the nature of the normal curve was briefly outlined. The standard instructions that were achieved both initially and when queries arose proved to be one of the particular advantages of the collective meetings.

When all assessments were complete, the purpose of the investigation was revealed, and there followed a short general discussion of the problem. " Background forms " asking for the necessary information about journeys and other details were distributed, and returned later by post.

The Rating Board.—The apparatus, of which thirty sets were used simultaneously, is illustrated in Fig. 1. It consists of a board, 16-in. square, on which is affixed a sheet of paper bearing 20 horizontal scale lines, each graduated and numbered from 0 to 16. These lines are red, but they are cut by vertical black lines at the 0 4 8 12 and 16 interval points. The vertical lines are made continuous with the ' trait headings ' which are reproduced on strips of paper which rest against a support at the head of the board so that they can be clearly seen.

The names, together with a code number for each child, are written upon small squares of card (½-in.×¾-in.) and these are inserted in ' clip stands.' These are small lead castings (¼-in. in diameter) with a pointer in front and a

---

* In these cases assessments were often the synthesis of two judgments. This had been anticipated by the writer with some misgivings, though considerations of etiquette had made it unavoidable. However, it must be remembered that in all schools of this kind, the head teacher is in intimate contact with the children and actual observation of the rating process in action suggested that the accuracy of judgments made in this way might well have been superior, particularly in reducing ' halo.' There was a continuous evaluative discussion when multiple judgments were made.

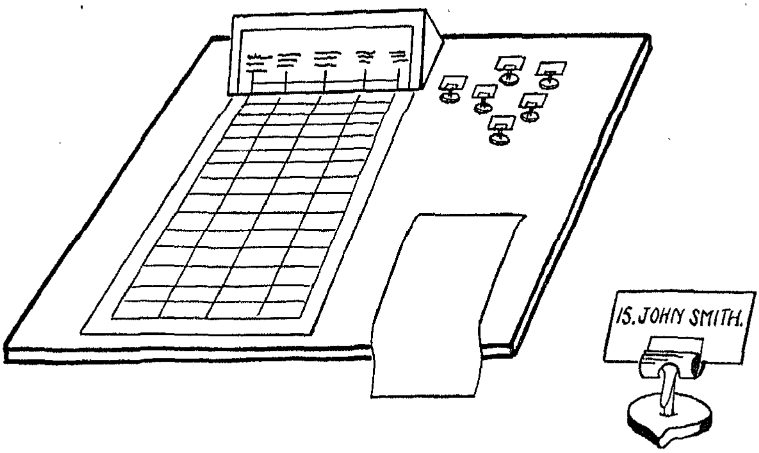


FIG. 1
The Rating Board.

very small bulldog clip cast into the top. The object of the clip stand is to combine manoeuvrability and stability with good display qualities.

The use of the board proceeds as follows : The clip stands are placed on the right-hand side and the first ' child ' picked up and placed upon the topmost red horizontal line in the position considered appropriate by reference to the trait headings at the top of the board. This is repeated with each child, one being placed on each of the horizontal lines, until the whole sample is arranged. When some or all of the clip stands are arranged on the board, an oblique view gives a simultaneous ranking of the children on the scale in relation to each other—and from appraisal of this, further readjustments are made. When the procedure is complete, the scale value on the red line nearest to the point of each child's clip stand is recorded on a form, against the child's code number. After the results have been ' taken off ' in this way, all the clip stands are returned to a random group on the right of the board, the trait headings are changed by removing the top sheet from the set—and the process is repeated, the ' children ' being replaced on the board according to the new trait headings.

The Behaviour Traits.—Scales for eleven behaviour traits were used. The trait headings (Appendix) were composed specifically for the purpose, with some help from a study by Cummings (1944) of the incidence of emotional symptoms in young children. The aim was to devise unidimensional scales, along a continuum from normal to neurotic, using symptoms that would be readily observable by teachers. The position of the neurotic end of the scale was alternated to reduce ' halo.'

The use of scales of this kind is an imperfect but necessary expedient in the present state of our knowledge. The only evidence for unidimensionality derives from those who make the judgments, and it should be reported that very little difficulty or confusion arose, except to some extent with Trait G—Withdrawal. The fact that teachers accepted the rationality of the scales is not, of course, in itself conclusive. It may be that the dimensions are stereotypes and that teachers fit their observations into them rather than abstracting ordered dimensions from their observations. At the same time, it must be recognised that scales of this kind do represent the synthesised observations of many students of human behaviour over the recent past.

IV.—RESULTS.

Reliability of Assessments.—Practical difficulties render the estimation of reliability, in the type of approach described, almost impossible. Assessments could be gained only from one judge or from a pair who unavoidably (cf. footnote, p. 103) worked together. Later reassessment would have been vitiated either by considerable transfer, after a short interval, or by shifts in the behaviour patterns, over a longer interval.

It was accidental that limited data became available from a school where twenty children were rated independently by a head teacher and by her assistant. Correlations were run for all behaviour traits. In all cases except C—Aggression (·449) the coefficient exceeds ·60, and for the combined index BDEFGI, it is ·775. The closest agreement is for I—Intelligence (·779).

Results for each Trait.—Trait J—Eating Difficulties, was found to be unsatisfactory because clear instructions were not given about the rating of those children who do not eat school dinners.

All other traits are considered in the analysis. Traits A—Absenteeism and H—Depression, are accorded distinct treatment because their distributions are markedly skewed. This also applies to Trait C—Aggression, but for the different reason that both extremes of the scale may be considered symptomatic of maladjustment. All traits with a normal distribution had s.d's between 3·855 and 4·100 except F—Popularity (3·424) which was, therefore, converted to 4·00. Slight skewness in I—Response to Affection, was corrected by transformation to a T scale and its deviations then contracted to the common mean value of 4·00. When necessary, scales were reversed so that a low value suggests poor and a high value good adjustment in all cases. The size and direction of differences on all the scales (except A, H and C) are, therefore, directly comparable statistically. It should be mentioned that this is not the same as asserting that they correspond to comparable degrees of maladjustment.

Fig. 2 shows the differences between the means for groups of children with short and long walking journeys and short and long transport journeys. (For Traits A and H, medians are used.) Separate means were computed for groups of children who have some distance to walk before meeting the school bus, etc., but the numbers are fairly small and since no consistent differences emerged, they have been combined with the ' transport only ' children, reference being made only to their transport time. Lines which slope downwards to the right provide evidence for the hypothesis that long school journeys are associated with maladjustment, and where the means for the short and long transport groups are below either or both of the means for walking—there is evidence that transport is associated with a further degree of maladjustment than walking.

The critical points for sub-dividing journey groups were decided from a combined index (see below). They are not necessarily the sub-divisions showing the maximum differences for any *individual* trait.

It can readily be seen that there is strong evidence for an association between the long school journey and maladjustment. The almost complete unanimity among the traits is perhaps more conclusive than the size of differences.

For the nine traits depicted, eight show more maladjustment for the long walking than the short walking group, and the ninth (E) contradicts the hypothesis to a negligible degree. The short transport group is also apparently more maladjusted than the short walking group for all traits except I, where the exception is again very small. The comparison between long transport and short transport accords with the hypothesis in eight out of nine of the traits, although

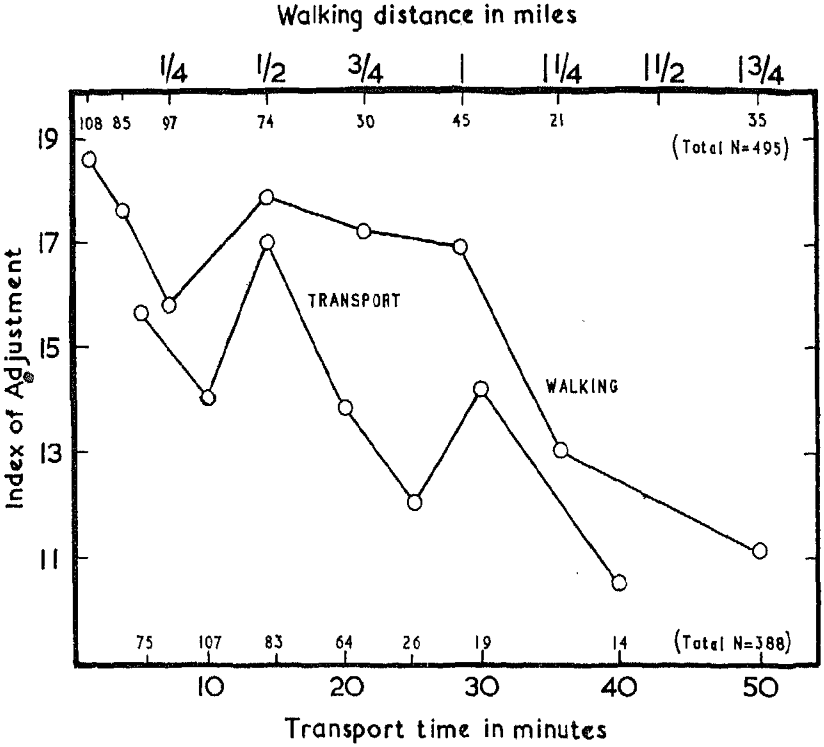


FIG. 2

Differences Between Means of Journey Groups for Each Trait.

P is for differences from short walking mean and is recorded on the right of means or halfway between two means when these have been combined. S.D. is approximately four for all traits except A and H, where the distribution is skewed. For these traits, P is for Chi^2 by the median test.

the single exception in this case (G) is the only one of sufficient size to command attention. It is not, however, significant at the 5 per cent. level and can hardly detract from the total picture.

The behaviour represented by Trait C—Aggression, is of considerable importance in any syndrome of maladjustment and it is unfortunate that the nature of the scale should make it less straightforward as a test of the hypothesis. Examination of the trait headings suggests that some maladjusted children of the withdrawn type might have been rated at the unaggressive end of the scale. On the other hand, the distinction must be recognised between a child who is excessively introverted and withdrawn and one who is 'calm and retiring.' If, notwithstanding, this end of the scale is interpreted as neurotic, the results contradict the hypothesis. If it is not, and the whole scale is interpreted as a dimension with increasing aggression equated with maladjustment, then significant differences (chi-square) in support of the hypothesis emerge between the two walking groups ($p < \cdot 05$) ; between the two transport groups ($p < \cdot 02$) ; and between the short walk and long transport groups ($p < \cdot 05$). If judgment is suspended on the unaggressive extreme of the scale, and only data from ' rarely aggressive ' onwards is considered, all the relationships mentioned above still support the hypothesis but only at the 10 per cent. level of confidence.

The inclusion of intelligence as a trait symptomatic of adjustment may require some explanation. Firstly, it is measured by teachers' observations of *performance* and this is the important consideration in the present context. Secondly, it would be difficult to counter-claim that the results are due to genetic differences between those who live in villages and those who live in isolated homesteads, because the performance of short transport children is higher than that of long walking children.

The ' unanimity ' to which we have referred is, of course, less decisive than would be the case if assessments for each trait had been made for separate and independent samples. As it is, we know that correlation between the measurements on such scales, due to syndromes of behaviour (with also, perhaps, a contribution from ' halo ' effect in the rating process) is more than likely. On the other hand if the traits cover a sufficient area of personality to involve more than one syndrome, as for example, hysteric and dysthymic, some of these correlations would be very low. The considerable task of computing intercorrelations between measurements on eight scales with a sample of 883 children has not yet been undertaken but would seem to be a logical step. In the present state of our knowledge, the empirical method of factor analysis would appear to be the best means of assessing the relative ' seriousness ' (factor saturation) of different symptoms.

Until this step is taken there are no valid grounds for computing a *weighted* combination of the scores of each child on the various traits.

This having been said, however, there is good justification for using at this preliminary stage a combined score in which the various traits are given equal weights. The underlying assumption is that the more symptoms of maladjustment that are exhibited, the more ' serious ' the condition is likely to be. It involves error because it assumes an equal ' seriousness ' for all symptoms. The use of an equally weighted combined scale is a *pro-tem.* device—but one which makes possible a number of fruitful explorations.

The Combined Index.—The combined index is based upon the six scales (B, D, E, F, G, I) with normal distributions and equalised s.d's. A and H, although discriminating highly between journey groups, could not be included. Trait G was included although in respect of one relationship (S.T./L.T.), it weakens the discriminating power of the index.

TABLE 1

COMBINED INDEX: MEANS FOR CHILDREN GROUPED BY SCHOOL JOURNEYS.

		N	Means	
Walking :	1 mile and less	439	17·403 ⇇ ⇇	⇇
	More than 1 mile	56	12·036 ⇇	16·796 ⇇
Transport :	15 minutes and less	265	15·502	
	More than 15 minutes.	123	13·203 ⇇	14·773 ⇇ ⇇

NOTE.—A low value implies maladjustment. Significant differences are indicated by brackets. Single arrows indicate significance at the 5 per cent. level, double arrows at the 1 per cent. level.

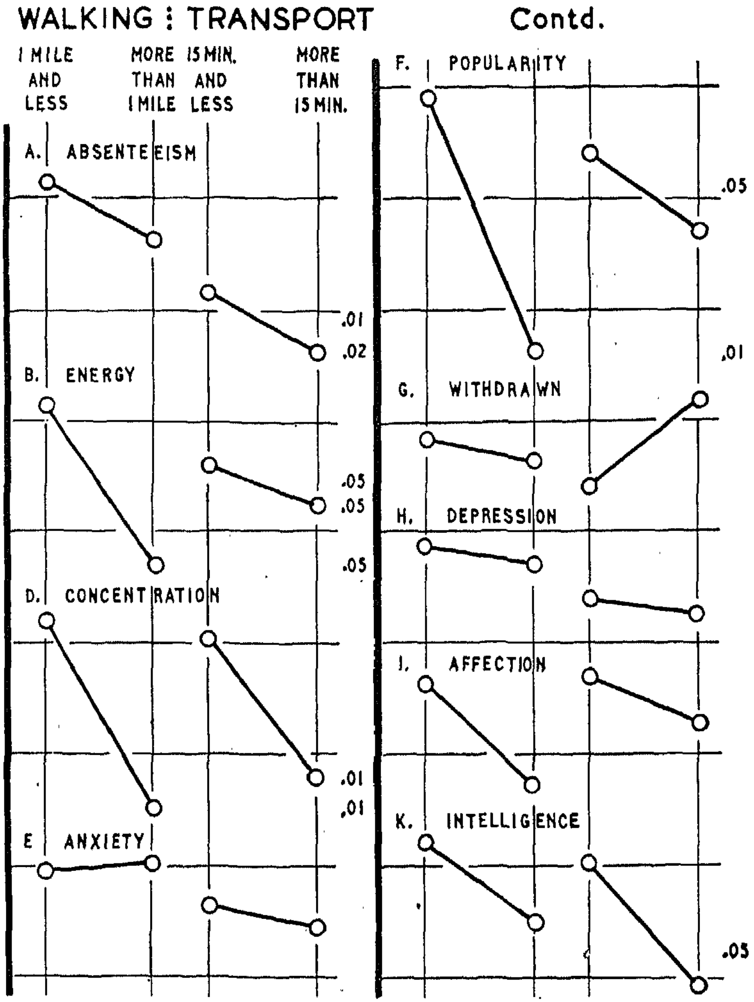


FIG. 3
Mean Adjustment (Combined Index) for Walking and Transport Groups by Distance and Time.

The extra stability afforded by the combined scale made it possible to plot the means for continuous sub-divisions of the journey groups and these are shown in Fig. 3. It will be seen that the sub-divisions which were selected to demonstrate the differences for individual traits are parts of what is probably a linear regression. The procedure of testing the null hypothesis for differences between means has, however, to be resorted to because of a J type distribution of journeys. Significance data for the combined index are shown in Table 1. (The addition of 20 units has been made to the means in all references to the combined index, in order to remove minus signs).

V.—SOME EXPLANATORY HYPOTHESES.

Three alternative hypotheses are now offered tentatively, to explain the relationship. They seem to be the most plausible, rather than the only ones possible.

(1) Children who live remote from village communities are less well adjusted because they come from families in which either the methods of child rearing are different; early social experience with other children is lacking; or the transmission of hereditary pre-disposition to neurotic disorders is more common because of a difference in the genetic stock. The children from such families have long school journeys, but this is merely a concomitant.

(2) Infant children who have long school journeys are physically fatigued at school.

(3) Infant children who have long school journeys are less well adjusted because they do not have adequate contact with their mothers, whom they *perceive* to be inaccessible for long periods.

Family difference hypothesis.—Although there is no reliable evidence to suggest that atypical families choose, for economic or other reasons, to settle away from centres of population, it is certainly possible and would account for our results. The hypothesis, however, is amenable to test and this fails to lend it any support.

Teachers were asked to classify children according to their residence, as market town, village, hamlet or solitary, each category being appropriately defined.

When town children are combined with village children (from whom they do not appear to differ), the numbers are large enough to ensure a fair reliability in the means of the combined index for each group, sub-divided by journeys. The general picture (Lee, 1956) shows the relationship between long journeys and maladjustment to be very little affected when the type of community is held constant in this way.

Before attempting to evaluate the two remaining explanatory hypotheses, it is necessary to examine the differences with sex and age, which contribute to an understanding.

Sex.—Both sex and age are important factors, and fortunately, the attribute and variable respectively are divided with rough equality throughout the journey groups and each does not, therefore, need to be held constant while the other is considered.

TABLE 2

MEANS FOR CHILDREN GROUPED BY SCHOOL JOURNEY AND BY SEX (COMBINED INDEX).

	Boys		Girls	
Walking:				
1 mile and less	15·719	14·795 (n=263)	10·274	19·065 (n=232)
More than 1 mile	8·125		17·250	
Transport:				
15 minutes and less	13·709	12·686 (n=204)	17·336	17·087 (n=184)
More than 15 minutes	10·729		16·472	
Total group		13·874 (n=467)		18·190 (n=416)

NOTE.—The differences between boys and girls for (*a*) walking and (*b*) transport are significant at the 1 per cent. level.

present study suggests that some of the groups may be relatively more deprived than others. However, within limits that are unlikely to be important, they all undergo the same periods of separation from home. It is suggested that this is equivalent to saying that they all undergo the same degree of deprivation only if we ignore a further crucial dimension of separation. This is *the child's perception of his accessibility to the mother.* For a fuller understanding, this needs to be added as a fourth to the set of variables relating separation and deprivation.

If maternal separation were the explanation for our results, the influence of school journeys would appear most comprehensible on a scale not of time, or of distance but of accessibility, as perceived by the child, to the home. Walking within the home village is a small barrier, walking in lonely lanes is a larger one, and transport is a serious obstacle. Transport to the rural child at school must mean the school bus, and this is not a line of communication that is always open. From the time of arrival to the time at which he is due to leave, and child time is very long indeed, the psychological barriers are up.

The fact that poor adjustment is more markedly associated with the long journey among boys fits more readily into this hypothesis than into one of fatigue. Although we lack any definite evidence that mothers are less ready to relinquish the dependent relationship with boys, such a conclusion is suggested by the data showing their generally retarded adjustment, and would seem a natural corollary of the higher value placed upon the male child in rural communities. Given this greater initial dependence (i.e., retarded phase of development), maternal separation would be expected to have more serious consequences for boys.

The age differences also fit an explanation in terms of accessibility. As the child grows older, his perceptions of space change as, and partly because, he gains physical strength. The longer walk would thus be expected to diminish as a barrier, but the unavailability of transport would remain a barrier as apparently insurmountable as before.

* * * *

ACKNOWLEDGEMENTS.—Grateful acknowledgement is made of the help and co-operation freely given by Mr. W. E. Philip, Chief Education Officer for Devon and his staff; and by the many teachers who took part. Mrs. Anne Ferris contributed invaluable clerical assistance.

VI.—APPENDIX.

A—ABSENTEEISM

Protracted spells with little apparent physical cause	Frequent short intervals	Occasional short intervals	Very Occasional short intervals	Never

B—ENERGY

Listless, very little energy	Tends to get tired	Quite energetic	Extra-energetic	Hyperactive, boundless energy, never tires

C—AGGRESSIVENESS

Calm and retiring, avoids provoking others	Rarely aggressive	Sometimes aggressive if provoked	Tends to enjoy being spiteful, belligerent	Tantrums, hysterical outbursts if frustrated

D—CONCENTRATION

High powers of concentration and persistence	Concentrates well on any task	Concentrates quite well if interested	Finds it hard to concentrate	Always restless, fidgetting, not attending

E—GENERAL ANXIETY

Fearful, anxious, timid	Sometimes fearful, anxious, timid	Composed and assured most of the time	Very rarely worries about anything	Absolute self-confidence and assurance

F—POPULARITY WITH OTHER CHILDREN

Extremely popular	Well liked by most children	Liked by most children	Rather disliked by some children	Extremely unpopular

G—WITHDRAWN

Quiet, withdrawn, day-dreams a lot		Only occasionally dreamy		Noisy, never seems to reflect at all

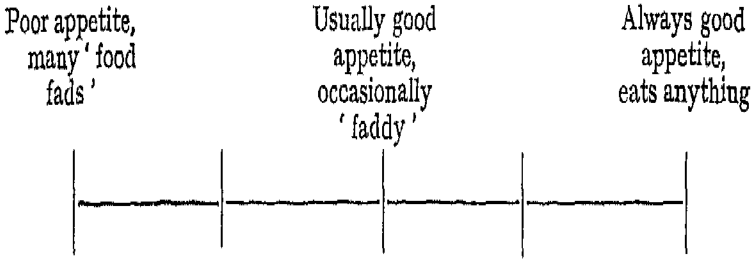
H—DEPRESSION

Always seems depressed, miserable	Often	Sometimes seems depressed, miserable	Very occasionally	Never seems depressed, miserable

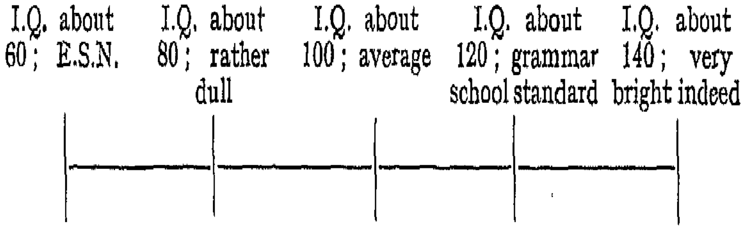
I—RESPONSE TO AFFECTION

Always responds warmly and unselfconsciously to attention and affection		Seems to appreciate attention and affection		Diffident, offhand repels every approach

J—EATING DIFFICULTIES


Poor appetite, many ' food fads '
Usually good appetite, occasionally ' faddy '
Always good appetite, eats anything

K—INTELLIGENCE


I.Q. about 60 ; E.S.N.
I.Q. about 80 ; rather dull
I.Q. about 100 ; average
I.Q. about 120 ; grammar school standard
I.Q. about 140 ; very bright indeed

VII.—REFERENCES.

BOWLBY, J. (1951). *Maternal Care and Mental Health.* Geneva : World Health Organisation.

CUMMINGS, J. D. (1944). The incidence of emotional symptoms in school children. *Brit. J. Educ. Psychol.,* XIV, 151-61.

LEE, T. R. (1956). *On the Relation between the School Journey and Maladjustment in Rural Infant Children.* Research Bull. I. Institute of Education, University of Exeter (mimeographed).

(*Manuscript received 29th June,* 1956.)

AN ANALYSIS OF THE VIEWS OF SUPERVISORS ON THE ATTRIBUTES OF SUCCESSFUL GRADUATE STUDENT TEACHERS

By J. D. C. ROBERTSON
(*National Foundation for Educational Research in England and Wales*).

SUMMARY. 1.—Eighteen supervisors of graduate students ranked fifty student attributes associated with successful practice teaching. The rankings were correlated and the matrix of correlations was factorised by the technique of Correlation of Persons.

2.—The lines of reference indicated among the participating supervisors showed:

(*a*) That there was limited agreement among them, amounting to 39 per cent. of the variance of their rankings, and

(*b*) That there was a number of subsidiary viewpoints among the supervisors, accounting for at least 17 per cent. of the variance. Groups holding these viewpoints appeared to vary in the emphasis placed upon the cognitive, affective (or personal) and conative aspects of teaching ability, and each group was divided again into:

(i) Those whose view was complex and analytical, concerned with adaptive aspects of personality, and

(ii) Those whose view was simpler and more intuitive, concerned with expressive aspects.

3.—Weighted average rankings of the attributes were obtained, and provide material for further research into an agreed scheme or system of student attributes. Groupings of the attributes made by the supervisors assisted in the interpretation of the factors, and were comparable in some respects with current theories of personality structure.

4.—The findings supported the conclusion of earlier researches that there is a need for an agreed scheme or system in the assessment of students' practice teaching. The limited amount of agreement found among the supervisors indicated a weakness in researches which have used uncontrolled assessments as criteria of teaching ability.

5.—The attribute group—Attitude and Insight in Dealing with Others was found, from the rankings of this group of eighteen experts, to represent for them the most important aspects of teaching ability.

I.—INTRODUCTION.

THE investigation to be described concerned the guidance of the teaching practice of graduate students by supervisors or tutors. Such guidance includes the following steps: (i) observation and the collection of data about a student; (ii) the evaluation of this data; (iii) diagnosis of the student's strengths and weaknesses; (iv) the planning and conduct of further practice; and (v) a final assessment. Each of these steps is modified by the supervisor's views on the objectives of his work, for during his experience he has developed systems of mental constructs, concepts or schemata related to teaching. It was upon these systems that attention was concentrated, and the investigation thus became a study of factors which modify the perceptions of supervisors. That is, supervision was considered as an interaction between the supervisor's view and the environment, here, the student in the practice situation.

To find how much agreement there was among the views of supervisors was the main object of the enquiry. It seemed that agreement was necessary if a number of difficulties were to be avoided; for example, students needed a clear idea of what was expected of them in their training, supervisors would find it helpful to know what attributes in students were being fostered by their colleagues, and, a point of great importance, assessments made by different supervisors might not be considered generally valid unless all were evaluating the same attributes. Further, if valid assessments of teaching ability are not available, research into the selection, training and appointment of teachers will be handicapped by the lack of a criterion.

Under the influence of the theories of the Gestalt and conditioning schools, there has recently been an emphasis on the learned or developmental aspects of perception. This trend would suggest that supervisors observe the work of students in ways which are likely to be idiosyncratic to a considerable extent, though there may well be a common basis due to the similarity of their experiences. There are three phases in perception in which supervisors may have developed differences: first, in the expectations they have about teaching and about students; second, in the characteristics of their perceptual activity, the ways in which they organize a configuration, or a system of relations in the field perceived; and third, in the amount of 'checking' they carry out to control the validity of their perceptions (Piaget 1950, Bruner 1951).

Piaget, in his analysis of perception, points out that perceptual activity, which involves centring upon different elements of the objective stimulus, may vary the relations perceived, and increase the relative value of those elements. If a supervisor should centre his observation upon a restricted range of student attributes, he might give them an importance which would seem excessive to some of his colleagues. It seems desirable, therefore, that supervisors should exchange interpretations, to develop in their perceptual activity an adequate, agreed range of centrings, and to make finer differentiations in their expectations. This would require a generally accepted conceptual system related to practice teaching. It was hoped that, if the analysis made here indicated a basis of agreement among supervisors, it would be a useful step towards meeting this requirement. Such a framework should direct attention to a more comprehensive range of attributes than the distortions and limitations of everyday perception of other personalities normally allow.

If a system already existed in the views of supervisors, controlling their perceptions, it should not have been difficult to show that there was general agreement about the student attributes which are associated with success in training courses. If, as expected in the light of perceptual theory, substantial agreement could not be established, and there was no evidence of a clearly defined system in the views of a group of supervisors, it might yet have been possible to find a number of respects, representing common elements in their experience, in which they did agree. Using these as a basis, a system acceptable to all might then be proposed.

It was also of considerable interest to know, if there was disagreement, whether this was due to views common to smaller groups, or whether it was so general that there was, in effect, no agreement. The hypotheses of the inquiry were based on these considerations.

II.—RESEARCH PREVIOUSLY REPORTED.

A study of previous research in teacher training revealed that a great deal of emphasis had been placed upon the assessment of students, most commonly by means of supervisors' ratings. Lins (1946) found these ratings to be the most reliable of the forms of assessment used in the Wisconsin studies, but, as there has been no satisfactory published evidence that supervisors agreed on the attributes of successful students, the validity of their ratings as criteria of teaching practice must be considered doubtful., Assessment has generally been based on the individual view held by a supervisor, though sometimes it has been derived from rating scales. These, too, have been found of doubtful validity by Lins and La Duke (1945), and they have reflected the views of those who constructed them. Barr (1945), noting that the supervisor's rating was the most reliable of his measures, stated that it would be difficult "to know what qualities the supervisors who rated these teachers had in mind " and concluded " . . . personality would seem to need a frame of reference."

In 1931, R. B. Cattell's survey of professional opinion on the qualities of a good teacher aimed at an explicit statement of the basis on which teaching ability was assessed. He sought a " clear, sound basis for discussion," a formulation of a " cumulative, intuitive opinion which is constantly being revised and tested in practical situations." Differences among experts, he believed, would be non-systematic, and eliminated by averaging a large number of opinions. He discovered discrepancies among the points of view of different groups, including head masters, training college staffs and inspectors. The present enquiry concentrates upon one group of experts in their own field.

For some time after Cattell's work research was concerned mainly with problems of selection and the pre-requisites for training, and various factors were related to ' teaching success ' ; but here, too, evidence for the validity of this criterion was inconclusive (e.g., Pinsent 1933, Turnbull 1934). Vernon (1950) classified the abilities of students by factor analysis and found they tended to fall into three groups : scientific, humanistic and practical. This classification has an interesting resemblance to that derived from supervisors' groupings of attributes in this investigation. Fuller references to the literature will be found in Barr (1945-48) and Evans (1951).

The survey outlined above revealed that research had had little success in a number of important respects : first, a standardised, objective means of assessment had not been established second, although supervisors' ratings appeared to be the most reliable form of assessment, there had been evidence of lack of agreement among them ; third, while these difficulties remained unsolved by the establishment of a valid criterion of teaching ability, other studies in the field were handicapped ; fourth, there had been many attempts to classify the attributes of teachers, and these, though they provided descriptive frameworks, had not been shown to assist in guidance and diagnosis by reflecting the actual relations of attributes in the functioning of personality In the present enquiry it was hoped first to find a basis of agreement which would provide a criterion, as a step towards the solution of these problems, and second, to confirm that without such a basis supervisors' assessments are of doubtful validity.

III.—PROCEDURE IN THE INVESTIGATION.

These objectives are not dissimilar to those of Cattell's investigation in 1931, but since then there have been considerable advances in the understanding of personality and in the use of statistical methods of classification. Advantage was taken of these in an attempt to ' map out ' the cognitive structures of a group of experienced, expert supervisors, who were working together in one of the larger university departments. The first step was an exploratory enquiry among some of the supervisors, who were asked which attributes of students they thought most helpful to success in practice teaching. This request was made informally, usually when accompanying the supervisor on a visit to a student.

When a collection of opinions was in hand, a list of fifty attributes was selected, arbitrarily, to make an adequate sampling with as little overlap as possible between items. Attributes like 'health' and 'intelligence' were excluded, as they were covered in other more specific items, and intending teachers are already highly selected in these respects. To establish the meanings of the items formal definitions were not used, but brief descriptions of kinds of behaviour or observable effects associated with the attributes were appended. A complete list of the attributes with the brief descriptions will be found in the Appendix. The additional description was often given in the words used by supervisors, and provided a graphic indication of what was intended. This method was used to save supervisors' time; being busy people, they were believed unlikely to be willing to attend to lengthy or abstract definitions. For the same reason it was not possible to ask them to rate students on such a large number of attributes to provide data for a conventional factor analysis. The items and explanations were arranged in random order and reproduced on a number of sets of cards, one item per card.

Supervisors were asked: (*a*) to re-arrange one of the sets of fifty cards in order, according to their views on the importance of the items in contributing to a high mark for practice teaching. They were to record their ranking, and then, (*b*) to sort the cards into bundles of attributes they considered similar. Eighteen supervisors completed the assignment, some independently, and some in the presence of the investigator, explaining their views.

IV.—TREATMENT OF THE DATA.

(*a*) *The rankings of the attributes.*—The eighteen rankings of the fifty items were tabulated and averaged. Then each ranking was correlated with the other seventeen, using Spearman's rank correlation formula, a procedure described by Burt (1940) as "Correlation of Persons." By this method it was possible to show the amount of resemblance between the views of any two supervisors on the list of attributes (Thomson, 1939) and to obtain a matrix of correlations which could be factorized. The highest correlation was between supervisors 9 and 17 ($\rho = \cdot 730$) and the lowest between supervisors 22 and 25 ($\rho = -\cdot 164$). Three-quarters of the correlations were found to be significant at the 5 per cent. level of confidence, and 61 per cent. were significant at the 1 per cent. level. The coefficient of concordance (Kendall, 1948), a measure of the communality of judgments for a number of rankers, was ·377 (significant above the 1 per cent. level).

(*b*) *The Factorial Analysis.*—It was expected that a factorial analysis would first indicate a general factor, the loadings of which would show how closely the supervisors conformed to a type, tending to rank the attributes in the same way. Further factors might be extracted, and each would group supervisors into sub-types with special points of view. Each supervisor would probably have also his own particular view, or specific factor. The matrix of correlations was factorized by the centroid method to provide "lines or terms of reference" (Burt, 1940) to assist in describing the views of the eighteen supervisors.

The statistical procedure was continued until a total of six factors had been extracted, when the variance of each factor and the communality of each supervisor was calculated. A check of the communalities estimated in the early stages of the procedure revealed some discrepancies, mainly due to the relatively high loadings on later factors of supervisors who had smaller loadings on the first factor. Accordingly, the communalities were revised and the whole analysis repeated. The loadings of the second and third factors were then plotted on a graph, but inspection of the result suggested that no profitable rotation of axes to achieve 'simple structure' was likely.

TABLE 1

FACTOR LOADINGS.

Supervisor No.	Factor No. 1	2	3	4	5	6	Communality
1	·706	+·279	+·226	—·151	—·152	—·154	·697
2	·584	—·528	—·183	+·343	—·092	+·169	·808
3	·537	+·162	—·258	—·250	+·224	+·184	·527
4	·567	—·204	+·149	+·186	+·176	+·015	·450
5	·569	—·559	—·100	—·339	+·256	+·170	·855
6	·809	+·095	+·165	—·241	—·081	+·071	·760
7	·119	—·322	—·240	+·050	—·204	—·188	·254
8	·760	+·021	—·046	+·243	—·121	+·190	·690
9	·741	+·340	+·159	+·121	—·091	+·202	·754
11	·586	+·379	—·123	—·231	+·120	—·195	·607
12	·570	+·174	+·245	+·194	+·140	—·240	·530
16	·507	—·218	—·168	—·244	—·258	—·146	·480
17	·779	+·304	+·058	—·028	—·106	+·123	·729
19	·699	+·215	—·247	+·216	—·239	—·256	·765
21	·730	+·186	+·266	—·083	+·020	+·002	·646
22	·118	+·059	—·435	+·213	+·328	+·163	·385
23	·614	—·033	+·263	+·107	+·158	+·029	·485
25	·713	—·353	+·281	—·093	—·007	—·056	·724
Total Variance ..	7·02	1·49	·871	·756	·549	·460	11·146
Average variance per cent.	39	8·4	4·8	4·2	3·1	2·6	61·9

The six factors accounted for 61·9 per cent. of the variance of the supervisors' rankings, so that 38·1 per cent. was due to individual views and to the error inherent in the procedure. Table 1 shows the Factor Loadings, with the total variance and average variance per cent. of each factor, and the communality of each supervisor. The first or general factor accounted for 39 per cent. of the variance, and the remaining factors, all of them bi-polar, accounted for 8·3, 4·8, 4·2, 3·1, and 2·6 per cent. of the variance respectively. The strictest tests (Guilford and Lacey ; Burt and Banks) suggested that only the first two factors were significant, but another (Ledyard Tucker) added the third and fourth. Guilford (1952) urges that as many factors as possible should be extracted, and the later factors here, though they contributed little to the total variance, indicated minor points of view which influenced a few supervisors' rankings.

(*c*) *The Groupings of the Attributes.*—It was possible to arrange the fifty items into a smaller number of groups by collating the supervisors' groupings of similar items in a table showing the frequency with which each attribute was associated with any other. Those pairs most frequently associated were selected and formed the basis of 'clusters' ; others were grouped with them, according to the frequency with which they had been associated, though towards the end of the classification the decisions had to be rather arbitrary. Eight clusters or groups could be distinguished, and each was provided with a general term to classify the items included. The groups were (with examples of the attributes in them) :

(1) Attitudes as a teacher (sincerity, enthusiasm for subject, etc.).
(2) Range of personality (experience, multiplicity of interests).
(3) Attitude and insight in learning to teach (grasp of principles and aims of subject, ability to discriminate, originality, etc.).
(4) Practical abilities in teaching (ability to organize, ability to finish a lesson, etc.).
(5) Teaching abilities (ability to interest, improves rapidly in skill, etc.).
(6) Personality qualities ; temperament and bearing (equable temperament vitality, attractiveness, etc.).
(7) Attitude and insight in dealing with others (good relations with other people, interest in people, etc.).
(8) Physical qualities (good appearance, physical stamina).

The average rankings of the items in each of the these groups are shown in Table 2. It seemed preferable to consider the findings of the enquiry mainly in relation to the groups of attributes, as they presented a simpler picture of the supervisors' rankings than did the more numerous specific items.

TABLE 2

GROUPS OF ATTRIBUTES.

GROUP	Average Rank	Wtd. Average Rank
1. Attitudes as a Teacher	27·3	31·3
2. Range of Personality	25·7	30·4
3. Attitude and Insight in Learning to Teach	20·9	20·9
4. Practical Abilities in Teaching	32·3	38·0
5. Teaching Abilities	24·5	28·4
6. Personality Qualities : Temperament and Bearing	27·5	33·0
7. Attitude and Insight in Dealing with Others	14·6	13·2
8. Physical Qualities	39·7	45·7

(*d*) *Weighting of Supervisors' Rankings.*—It was now possible to weight the supervisors' rankings on the basis of their loadings on the first, or general factor ; that is, according to the closeness of their agreement with the poiut of view common to the whole group. The method used is described by Thomson (1939) and provides, in the averaged, weighted rankings a hypothetical point of view closer than the original averages to the view which would be most representative of the whole group of supervisors. The changes in item ranking after weighting are reflected in the weighted averages of the attribute groups in Table 2. As the ranks of the other groups were reduced, the relative positions of the two groups originally ranked highest, 7 and 3, were improved.

V.—INTERPRETATION OF THE FACTORS.

It was possible to make a tentative interpretation of the factors on three bases :

(2) External evidence : the findings of other enquiries using the technique of correlating persons. In reports of enquiries by Dewar (1938), Eysenck (1940a, 1940b, 1942) and Peel (1945), two tendencies were noticeable : (*a*) the subsidiary factors appeared to indicate whether the rankers made a complex or simple approached to the stimuli presented, and (*b*) the first subsidiary factor tended to

be related to cognitive aspects of the stimuli, the second to affective (or personal) aspects.

(2) Internal evidence: variations in the rankings of groups of attributes, and of single attributes, by the cores of the groups of supervisors (i.e., those supervisors with the highest positive or negative loadings on each factor).

(3) Confirmatory evidence: comments made by some supervisors consulted informally about the groups among their colleagues.

The first or general factor.—Sixteen of the eighteen loadings on this factor were significant and indicated how closely each supervisor approached agreement with the viewpoint representative of the group. (It is not claimed that the group itself was representative of all supervisors, but it was reasonably large and varied. Most of those taking part were preparing students for grammar school teaching, but two were senior members of other departments). The seven with the highest loadings included men and women, seniors and juniors; they covered a wide range of years of experience, and represented a number of subject departments. Colleagues could not readily give a reply when asked what this group had in common. Their views are best represented by the weighted rankings shown in Table 2.

The second (first bi-polar) factor.—As suggested by other researches, this factor appeared to distinguish two groups of supervisors with special cognitive approaches to the ranking, the positive core more complex, the negative core simpler. The two cores disagreed most markedly about attribute groups 3—Attitude and insight in learning to teach (ranked 12·5 and 29·9) and 6—Personality qualities: temperament and bearing (30·8 vs. 23). The positive core emphasised adaptive aspects of personality. The negative core, on the other hand, ranked items 32—" Ability to make lessons enjoyable," and 42—"Attractiveness," higher, and generally valued items representing the " pupil's point of view," or the expressive aspects of personality. Colleagues commented on the positive core:

"A sound conception of teaching, look for a workmanlike approach."
" Sets a high standard, rather exacting."
" Proceed by analysis," etc.

And on the negative core:

" Would make a judgment on an impression of a person."
" Less discriminating."
" Intuitive apprehensions," etc.

The data available for interpretation from three sources were thus consistent, and supported the conclusions that this grouping was based on cognitive differences, and that two further quite distinctive points of view were influencing these supervisors in their rankings.

The third (second bi-polar) factor.—External evidence suggested that this factor would have an affective or ' personal relations ' basis. The positive core would have a complex view of this aspect of teaching, and the negative core a simpler view. This appeared to be so, as the most marked differences in their ranking concerned attribute group 7—Attitude and insight in dealing with others (7·5 vs. 27·9). Individual items about which they differed were 4—"Ability to organise," and 11—" Effectiveness in teaching " (favoured by the positive core) and 28—" Confidence of bearing," and 3—" Extraversion " (negative core). Colleagues commented on the positive core:

" Interested in the broader aspects of the education of children."
" Orderly, with executive powers."
" Good social relations, nice persons," etc.

And on the negative core:

" Like people without discrimination, friendly."
" Simple, Wordsworthian appreciation of homely things," etc.

Again, the data were consistent and indicated two more points of view contrasting complexity of response and penetration in personal relations with a ready response to expressive characteristics of people.

The fourth (third bi-polar) factor.—There was no external evidence to suggest the nature of this factor, except Spearman's well-known third division of aspects of the mind—the conative; and this basis was supported by the data. The positive core tended to value attribute group 2—Range of personality, while the negative core preferred group 4—Practical abilities. Differences in the ranking of items were found to contrast the attributes of a mature, poised, discriminating student with those of a robust, bright, enthusiastic student. Colleagues' comments suggested opposing views of the student in action, learning to teach. The contrast was between insight into rapid but unspectacular forms of adjustment on the one hand, and a preference for more tangible evidence of drive on the other.

The remaining factors will not be treated in detail, as they were not of great importance. Study of the fifth factor suggested contrasting views of the 'craft' of teaching: vigor, originality and discrimination *versus* careful, friendly effectiveness. In the sixth group the positive core favoured achievement and progress in teaching skill, while the negative core were content with adequacy in the classroom, provided it was combined with more general personality strength.

VI.—DISCUSSION.

The findings of the investigation provided consistent support for the hypotheses advanced. Among the supervisors who took part there was a measure of general agreement about the attributes of successful students, accounting for 39 per cent. of the total variance of their rankings. There was also evidence of the existence of at least three other groups (each divided into two sub-groups) with different points of view, accounting for a further 17 per cent. of the variance. (The first four factors accounted for over 70 per cent. of the factor content of the rankings of six supervisors.) The remaining variance was no doubt due to points of view peculiar to individual supervisors, as well as to the error inherent in the method of investigation.

The degree of general agreement about the attributes which contribute to success in practice teaching was not high. There were some different opinions common to groups of supervisors, and also some idiosyncrasy; and a number of supervisors showed only a limited agreement with any of their colleagues. In short, the supervisors' views of the material presented to them varied substantially, and there were clearly a number of different approaches to students' teaching practice. This evidence, like that reported by Barr and others, does not support the practice of awarding teaching marks or assessments without a valid scheme of reference.

Further, the variations in the views of supervisors made it clear that they do not proceed on the same lines or towards the same objectives in their guidance. As the classifications made in the enquiry were supported by the generalisations of colleagues, the tendencies shown by supervisors in their rankings seemed to persist in their perception as a whole, and in their appraisal of students. The differences revealed may well be due to professional and other experience, as suggested by recent theory. While it is not suggested that guidance could be standardised, an accepted system or theory like that sought by Cattell should facilitate communication among supervisors and stimulate their perceptual activity to a wider range of centrings. It would be of particular value in assessment, and in clarifying for students the objectives of their training. The development of the basis of agreement demonstrated here, in the weighted average rankings, might well provide a criterion of teaching ability and a partial solution to the problems revealed by previous research. An accepted criterion would open the way to valid assessments and fruitful research in selection, training methods and placement.

Although the amount of agreement among the supervisors was limited, it was sufficiently widespread and substantial to encourage further research aimed at a generally acceptable conceptual system related to the guidance of practice teaching. Like most theory, this system would not be considered final, but would be flexible and constantly developing. As the data obtained in this enquiry provided a useful starting point, work on this is proceeding.

The analysis of teaching ability which resulted from the investigation (Table 2) reflects, in an interesting if general way, the structure of personality as this has been represented in some current theory (e.g., Bronfenbrenner, 1951). In a preliminary classification of the attributes of students it gave recognition to the importance, first, of interpersonal relations, and second, of capacities of an intellectual kind. The emphasis placed by most groups of supervisors on the attribute group, Attitude and insight in dealing with others, is perhaps the most significant finding of the investigation. In the opinion of these experts, in the selection, training and assessment of students, attributes related to this grouping should be given the greatest weight.

* * * *

ACKNOWLEDGEMENTS.—The work described was carried out under the guidance of Professor P. E. Vernon, of the London University Institute of Education. His assistance, and that of the supervisors who provided the data, is gratefully acknowledged.

VII.—REFERENCES.

BARR, A. S. (1945 and 1948). Measurement of teaching ability. *J. Exp. Educ.*, 14, 1-5, 16, 203-283.

BRONFENBRENNER, U. (1951). *Towards an Integrated Theory of Personality* in BLAKE, R. R. and RAMSEY, G. V. *Perception—An Approach to Personality* (Symposium). New York: Ronald Press.

BRUNER, J. S. (1951). *Personality Dynamics and the Process of Perceiving* in BLAKE and RAMSEY, see above.

BURT, C. (1940). *The Factors of the Mind.* University of London Press.

CATTELL, R. B. (1931). The assessment of teaching ability. A survey of professional opinion on the qualities of a good teacher. *Brit. J. Educ. Psychol.*, 1, 48-71.

DEWAR, H. (1938). A comparison of tests of artistic appreciation. *Brit. J. Educ. Psychol.*, 8, 29-48.

EVANS, K. M. (1951). A critical survey of methods of assessing teaching ability. *Brit. J. Educ. Psychol.*, 21, 89-94.

EYSENCK, H. J. (1940 a). The general factor in aesthetic judgment. *Brit. J. Psychol.*, 31, 94-102.

—— (1940 b). Some factors in the appreciation of poetry. *Character and Pers.*, 9, 160-167.

—— (1942). The appreciation of humour: an experimental and theoretical study. *Brit. J. Psychol.*, 32, 295-309.

GUILFORD, J. P. (1952). When not to factor analyse. *Psychol. Bull.*, 49, 26-36.

KENDALL, M. G. (1948). *Rank Correlation Methods.* London: Chas. Griffin.

LA DUKE, C. V. (1945). Measurement of teaching ability. *J. Exp. Educ.*, 14, 75-100.

LINS, L. J. (1946). Prediction of teaching efficiency. *J. Exp. Educ.*, 15, 2-60.

PEEL, E. A. (1945). On identifying aesthetic types. *Brit. J. Psychol.*, 35, 61-69.

PIAGET, J. (1950). *The Psychology of Intelligence.* London: Routledge and Kegan Paul.

PINSENT, A. (1933). Pre-college teaching experience and other factors in the teaching success of university students. *Brit. J. Educ. Psych.*, 3, 109-125.
THOMSON, G. H. (1939). *The Factorial Analysis of Human Ability.* University of London Press.
TURNBULL, G. H. (1934). The influence of previous teaching experience on results obtained by students in a university department of education. *Brit. J. Educ. Psychol.*, 4, 1-9.
VERNON, P. E. (1950). *The Structure of Human Abilities.* London : Methuen.

VIII.—APPENDIX.

LIST OF ATTRIBUTES WITH ADDITIONAL EXPLANATIONS

	ATTRIBUTE.	ADDITIONAL EXPLANATION.
1.	Ability to interest	" Can hold a group " ; makes work meaningful.
2.	Practical ability	Able to handle equipment, material ; can improvise.
3.	Extraversion	" Controlled showmanship " ; " can put it over " ; " putting oneself across."
4.	Ability to organize	Able to imagine a lesson in the classroom ; prepares well thought out plans, or a purposive procedure.
5.	Equable temperament	Poised ; not easily put out ; stable.
6.	Vitality	Vigour of mind ; interest, enthusiasm, zest ; spontaneity.
7.	Good Relations with Other People	Sympathetic to pupils ; co-operative with staff, supervisors.
8.	Improves Rapidly in Skill	Has learned quickly to present his lesson, to develop learning.
9.	Awareness of Class	Teaches every individual ; is sensitive to response, to how class sees things.
10.	Good Appearance	" Good looks " ; " well dressed and groomed"; likely to appeal to young and enhance prestige.
11.	Effectiveness in Teaching	Pupils learn well ; successful lessons ; develops knowledge, skill, interests.
12.	Adequacy	Shows no strain or tension ; " seems able to cope."
13.	Sense of Proportion	" Able to laugh with a class (not at them) " ; humour ; common sense.
14.	Grasp of Principles and Aims of Subject	E.g., understands scientific method, or semantics, or has insight into artistic values, sees value of evidence, etc.
15.	Experience	Able to draw, in teaching, on a wide experience of life outside school.
16.	Social Acceptability to Pupils	Pleasant, agreeable, friendly, tactful, well-mannered, good-natured.
17.	Active Attitude in Learning to Teach	Willingness to experiment, to try another way, to use suggestions ; interested in solving problems of teaching.

18.	Awareness of Nature of Learning Process	Knows learning is dynamic, continuous.
19.	Punctuality, and Regularity of Attendance	" Is always on the job " ; " Doesn't make excuses."
20.	Insight into Techniques of Teaching	Has grasped suggested ways of teaching ; knows how and why methods demonstrated are to be used.
21.	Cheerfulness	Seems to enjoy work ; is bright and buoyant ; " a happy disposition."
22.	Ability to Criticize One's Own Teaching	Self-critical ; " able to step outside oneself."
23.	Ability to Develop a Point	Makes use of answers ; keeps class thinking.
24.	Ability to Communicate	Able to use written and spoken English ; fluency.
25.	Flexible Attitude to the Theory of Teaching	Willingness to consider new ideas ; to accept more complex views, or " Gestalts."
26.	Persistence	Doesn't give up easily ; keeps on trying ; not deterred by difficulties.
27.	Multiplicity of Interests	Brings keen outside interests, or a variety of outside activities to bear in teaching.
28.	Confidence of Bearing	" Good Presence " ; faces class firmly and pleasantly ; assured.
29.	Ability to Use Aids	Uses blackboard effectively ; sketches ; makes good diagrams ; uses pictures, films, charts, etc.
30.	Imagination	Able to imagine pupil's views, his vocabulary, difficulties and reactions.
31.	Enthusiasm for Subject	Keen about subject ; feels its importance ; " can communicate a devotion."
32.	Ability to Make Lessons Enjoyable	Pupils find lessons fascinating ; they would like more.
33.	Ability to Finish a Lesson	Sums up well, and creates interest in the next step ; times steps well.
34.	Ability to Organise Class	Can supervise all activities ; give out equipment ; " manage."
35.	Interest in People	Knows the pupils ; seeks to understand and help them.
36.	Sincerity	" Sound ideals " ; seriousness ; recognizes value of teaching work.
37.	Reliability	May be depended upon ; will do what he can.
38.	Ability to Question Well	Is able to secure a good response ; asks useful and successful questions.
39.	Social Maturity	At ease with others ; not on the defensive, sentimental, self-important or authoritarian.
40.	Ability to Discriminate	Selects important things ; understands principles and works on that basis.

41.	Uses Methods Introduced in the Course	Tries methods explained or demonstrated; attempts to employ suggestions.
42.	Attractiveness	"Makes pupils desire to please."
43.	Adaptability	"Makes the best of the situation"; "can improvise or modify intentions."
44.	Tolerance	Accepts differences, immaturity; allows other opinions, not repressive.
45.	Self-sufficiency	Self-reliant; not always seeking help; independent.
46.	Adequacy of Speech	Clarity of speech; articulation; strength, variation in the use of the voice.
47.	Physical Stamina	Good physique, digestion, "nerves."
48.	Originality	Invents new approaches; sees subject and lesson from a new angle; "creative presentation."
49.	Diligence	Prepares regularly and adequately, attends to detail.
50.	Ability to Explain	Makes points clear; simplifies effectively; illustrates well.

(*Manuscript received* 29*th May*, 1956.)

EINSTELLUNG RIGIDITY, INTELLIGENCE AND TEACHING METHODS

By KENNETH M. MILLER
(*Department of Psychology, University College, London.*)*

SUMMARY.—In a study with 120 technical and secondary modern schools boys it has been shown that:
(i) There is a significant relation between Einstellung rigidity or mechanization in problem solving and intelligence, and
(ii) That when sub-samples of technical and modern school boys matched for intelligence, are compared, the Einstellung effect occurs significantly more often in the modern school group. This result is discussed as a function of teaching methods and attitude to school.

I—INTRODUCTION.

(i) *Recent research on rigidity.*—Papers in earlier numbers of this and other journals have reported the extensive research on perseveration carried out by Spearman, Stephenson, Cattell and other members of the London school. Recent American investigations into the generality of rigidity as measured in a variety of situations [2, 3, 4, 5, 7, 10], have either included tests of 'classical perseveration' or else assumed that rigidity, as demonstrated by the early tests of perserveration, was similar to rigidity measured by more recently developed techniques.

Of these, the techniques most frequently used were first reported by Luchins in 1942,[8] when describing the results of an investigation into the mechanization of problem solving. Of the several methods he described, the one involving a series of arithmetical problems (sometimes referred to as the water jar Einstellung test, the special properties of which are stated below) has been most extensively used in subsequent studies. In his investigation Luchins demonstrated that mechanization of problem solving occurred with subjects of all ages and educational levels. In other experiments he investigated the factors which might have been salient in the development of such mechanization. These included the use of special instructions, intelligence level of subjects, the relation of the experimenter with the subjects and school environment. For some of these factors the evidence was quite clear, while for others the findings were stated to be only tentative. This was particularly the case for school environment, methods of teaching and intelligence. For school environment the results of twenty-nine subjects from a progressive private elementary school were compared with the results of (American) public school subjects. Mechanization occurred but to an extent " on the whole somewhat less than in the public schools." In view of later work it is significant that the I.Q. range in the private schools was from 120-149, while the range in public schools was from 70-169.

Further consideration of the results from different public schools indicated that in a school using ' drill methods ' extensively, the Einstellung effects were greater than in a school making more use of ' progressive methods.' No precise test was carried out on these data nor on the data for intelligence. The relation between intelligence and mechanization of problem solving was considered by dividing subjects into three I.Q. groups—high (over 110), medium (90-110) and low (less than 90). No significant relation was found, but members of both high and low groups displayed more mechanization than the middle group. Rokeach later reported a negative but non-significant correlation between intelligence and rigidity.*

---

* Now at the University of Tasmania, Hobart, Australia.

(ii) *Aim of Investigation.*—Luchins, in presenting the above findings, suggested the need for further study of the relation between intelligence and mechanization of problem solving and of the influence of educational factors. The present study was designed to obtain further evidence about both of these aspects.†

II.—School Environment.

In his study, Luchins made no attempt to define educational environment, but rather left the reader to infer on the basis of previous knowledge, the nature of, and the differences in, environment likely to exist between private and public schools (and between different public schools). While variables such as size of class, socio-economic status of parents, progressive methods of teaching and physical conditions could have contributed to the results, so too could other factors such as the personality of the teachers, the attitude to school of the pupils, and teaching methods.

The present study was conducted in an integrated secondary-modern-technical school in central London, thereby minimizing the effect of some of the factors not readily controllable in the Luchins study (which used different schools) and providing the opportunity for a more precise investigation of other factors ; in particular, the effect of differences in teaching methods. Thus, such aspects of educational environment as physical surroundings, teacher personality, socio-economic status of parents,‡ were comparable for all boys, while other factors such as teaching methods differed. Another factor, less certainly an environmental one, but nevertheless, potentially important, was the attitude of the boys to school.

In the school a warm permissive atmosphere prevailed with the head and other teachers having the confidence of the boys, who frequently accepted the invitation of the head teacher to ' come and talk things over.' Nevertheless, the attitudes of the boys in the two streams did differ in that boys in the technical stream were keenly preparing for formal examinations and/or to enter apprenticeships, while those in the modern stream were without such clear objectives and in many cases would not have been regretful when schooling ceased.§ Although the same teachers taught boys in both streams, they did, according to their own reports, adopt different approaches in the two streams. In the modern forms 'drill methods ' were used to a greater extent than in the technical forms.

---

* In a critical review of a large number of studies in which the water jar Einstellung test had been used, Levitt[6a] indicates that several unpublished and published studies (including that of Rokeach) show a consistent low negative relationship with intelligence. Levitt also shows that the claim that the water jar test is a general measure of rigidity is open to doubt as only five of thirty-one studies report positive results. More importantly, he demonstrates that in those studies in which stress has resulted in increased rigidity, the results can be more readily interpreted in terms of learning theory. These and other points in the review, whilst important for the wider study of rigidity, do not bear directly on the present paper which is concerned with the relation between the test as a measure of mechanization in problem solving and educational factors.

† The study was also concerned with the generality of rigidity—the results of which will be reported separately.

‡ Occupations of parents with one or two exceptions, were rated 5 or 6 on the Hall Jones system of occupational classification.

§ A pleasing feature of the school is that any modern form boy who shows keenness is allowed to transfer to the technical stream.

III.—EXPERIMENTAL PROCEDURE.

(i) *Einstellung problems.*—The arithmetical problems referred to above are similar to the water jar problems of the Terman Merrill test in that the subject is required to indicate how to obtain certain exact quantities of liquid by using a number of containers and not resorting to guessing. The following standard description of a set of Einstellung problems indicates how the subjects are ' set ' to solve the problems in a certain way prior to the introduction of problems apparently similar but which may be solved by a shorter method. Notice is then taken of whether the subjects see the short method or continue to use the longer method.*

> " Problem : Given : an empty 21 quart jar, an empty 127 quart jar and an empty 3 quart jar ; measure 100 quarts of water. If the jars in the order presented in each statement of the problem are labelled A, B and C, respectively, then the method which solved the problems (2 through 6) can be formulated as B − A − 2C ; for example, 127 − 21 − 2 × 3 = 127 − 27 = 100. This method is known as the Einstellung or E method, and problems, 2 through 6, which are solvable by this method are known as Einstellung or E problems. Problems 7 through 11 are known as test problems, as they indicate the effect of the E method in the previous problems. Problems 7, 8, 10, 11, are solvable not only by the E method, but by more direct methods A − C or A + C, respectively. The ninth problem, however, cannot be solved by the Einstellung method, but is solvable by the A − C formula. This problem has been called an extinction task . . . In addition to serving as a factor to bring about recovery from the Einstellung, this ninth problem serves also as a crucial test of the Einstellung effect." (9, p. 89.)

It was decided to follow Luchins' procedure for administration by including an extinction problem between the two pairs of crucial problems rather than use an initial control problem. As Luchins[7] had also found that there was no diminution in the amount of E-effect when all methods of solution were demonstrated, the number of examples was increased to three, enabling all types of solution to be demonstrated. This meant that every subject was made aware of the short method of solution, and any failure to use this method could then be attributable to the development of set and its interaction with any other variable shown to be influential in the determination of rigidity of behaviour in this situation.

The problems were written one at a time on the blackboard, the subjects working the problems on a single sheet appropriately ruled, were allowed 2½ minutes for each problem. As the groups were small, it was possible for two people to exercise close supervision, and to attend to any queries. This method of presentation was adopted after careful consideration of methods used by other investigators. Rokeach[10] presented the problems individually and had the subjects work each problem on a separate page of an examination booklet; Luchins[8] presented the problems individually but did not have separate answer pages ; and Ainsworth[1] presented the problems printed on a single sheet. No other reports indicate clearly the method of presentation, the authors apparently considering it of no importance—in fact, Levitt and Zehlen[7] have specifically stated that method of presentation does not influence the Einstellung-effect.

In scoring the problems, it was considered advisable to retain the criterion of discarding cases with mistakes in the Einstellung problems, but inadvisable to discard those failing to solve the extinction problem. The score used was the number of extinction and control problems correctly solved by the direct method—giving a range of 0 to 5 with a low score indicating high rigidity. The actual problems used in the study were based on those of Luchins,[8] Rokeach[10] and Cowen[5] and were tried out with subjects of comparable age and intelligence in another London school: results indicated that the problems were arithmetically suitable for the present research. The problems are presented in the Appendix together with the instructions for administration.

---

* Following Luchins, the prefix ' E ' stands for Einstellung—e.g., E effects, E problems. Also Einstellung effect, rigidity and mechanization of problem solving are used interchangeably throughout.

(ii) *Subjects.*—One-hundred-and-twenty third and fourth form boys took part in the main study and of these seventy-three were in technical school forms and forty-seven in modern school forms.* In the technical form group, the age range was from 169-198 months with mean age of 180 months and S.D. 4·4 months. For the modern form group the range was from 166 to 182 months, mean 171 months and S.D. 4·8 months.

IV.—RESULTS.

(i) *The Relation between Intelligence and Mechanization in Problem Solving.*—With the limited score range of six points, rank correlation (corrected for ties) was used to test the relation between intelligence and mechanization or rigidity. The correlation was significant beyond the 1 per cent. level (rho= – ·304). This result then is more definite evidence of the trend noted by Rokeach who obtained a negative but non-significant product moment correlation.

A further test of the relation was made by sub-dividing the sample into high, medium and low I.Q. groups, using the same cutting points as Luchins and then applying a chi-square test. The detailed distribution of the sample by rigidity score and intelligence group is shown in Table 1. Inspection of this table supports the correlation analysis. As presented, the table does not meet the minimum requirements for the application of chi-square, so condensation was effected by categorising those subjects with 0, 1 and 2 problems solved by the direct method as ' rigid ' and those with three or more problems solved by this method as ' non-rigid.' For the data in this form the value of chi-square is significant beyond the 1 per cent. level. ($\chi^2=13\cdot1$, d.f. 2). Thus, contrary to Luchins' suggestion there is evidence of a significant negative relation between intelligence and rigidity.

TABLE 1

DISTRIBUTION OF SAMPLE BY INTELLIGENCE AND E-EFFECT SCORE.

Intelligence Group	Number of Problems solved by the Direct Method						
	0	1	2	3	4	5	Total
High (111+)	7	2	5	9	7	4	34
Medium (90-110)	11	17	16	11	10	5	70
Low (89—)	9	2	4	1	0	0	16
Total	27	21	25	21	17	9	120

(ii) *The relation between rigidity and educational factors.*—If the conclusions of Luchins concerning the relation of rigidity with intelligence had been supported by the above analysis, a direct comparison of the results of technical form and modern form subjects would have been appropriate. But, in view of the observed significant relation, some control of intelligence is required. When each of the school groups is broken down into High, Medium and Low intelligence sub-groups, most of the High group are, not unexpectedly, found to be technical stream pupils while most of the Low group are found to be modern stream pupils.

---

* Altogether 148 boys participated in the study, the results of twenty-eight having to be discarded because of arithmetical mistakes or absence from one or more tests.

TABLE 2

DISTRIBUTION OF SAMPLE BY INTELLIGENCE INTO "RIGID" AND "NON-RIGID" GROUPS ACCORDING TO NUMBER OF PROBLEMS SOLVED BY DIRECT METHOD.

Intelligence Group	Rigid 0—2	Non-Rigid 3—5
High	14 (41%)	20 (59%)
Medium	44 (63%)	26 (37%)
Low	15 (94%)	1 (6%)

In the middle group there were forty-one technical stream pupils (mean I.Q. 102·8) and twenty-nine modern stream pupils (mean I.Q. 100·6). A *t* test for unrelated means indicates that the difference between means is not significant (t=0·28). These sub-groups were categorised into rigid and non-rigid, as shown in Table 3. Applying chi-square to this table the result is significant beyond the 1 per cent. level (chi-square=6·5, d.f. 2). Further, when two groups matched for I.Q. are abstracted from the middle group and the test repeated, a result of the same order is obtained.

TABLE 3

DISTRIBUTION OF MIDDLE INTELLIGENCE GROUP SUBJECTS BY EDUCATIONAL GROUP INTO "RIGID" AND "NON-RIGID."

	Rigid	Non-Rigid	Total
Technical	20 (49%)	21 (51%)	41
Modern	24 (83%)	5 (17%)	29
Total	44	26	70

Finally, in investigating the difference in degree of mechanization between technical and modern stream subjects an analysis taking age into account was conducted. Although the correlation between age and rigidity is negative but non-significant, it was considered that the greater range and higher mean age of the technical group might have had some influence on the results. So, from the total sample, two groups (each of twenty-four cases) matched for age and intelligence, were selected as in Table 4. The application of chi-square again resulted in a value significant beyond the 1 per cent. level (chi-square=8·46, d.f. 1).

TABLE 4

DISTRIBUTION OF TECHNICAL AND MODERN GROUPS, MATCHED FOR AGE AND INTELLIGENCE, INTO "RIGID" AND "NON-RIGID."

	Rigid	Non-Rigid	Total
Technical	8	16	24
Modern	19	5	24
Total	.. 27	21	48

V.—DISCUSSION.

(i) *Einstellung rigidity and intelligence.*—In the present study with a sample of 120 cases, a significant relation between intelligence and Einstellung rigidity has been demonstrated, thus helping to clarify the position opened up by Luchins in 1942, and, with the exception of Rokeach, ignored by later investigators who have focussed their attention on Einstellung rigidity as a personality variable. The present writer contends that while personality is not unimportant in regard to Einstellung rigidity, intelligence plays a more important part in its development than has been previously acknowledged.

(ii) *Educational Factors.*—As the analyses have shown highly significant differences in Einstellung effect between technical stream and modern stream subjects matched for intelligence, one must consider whether other factors in the educational situation could have been influential in the development of Einstellung rigidity.

In view of the teacher reports on difference in approach to the teaching of the two types of form it seems reasonable to conclude that the drill or routine, methods which are used with the modern stream forms constitute an important factor. It seems likely that in being trained to utilise sure methods of work, the pupils tend to approach all new situations in this way, thus failing to show the same flexibility in attack as do other pupils in whose formal training emphasis is placed on finding alternative methods of solution. Depending on what one considers the aim of education, it could be argued that in suiting teaching methods to the less able modern stream pupils, the more able are prevented from developing as flexibly as they otherwise might have done. On the other hand, if it is considered desirable to train pupils to approach the solution of problems in a mechanical way, the above findings need cause no concern.

When considered in relation to recent reports by Vernon[11] and Daniels[6] concerning change in I.Q. of children subjected to streaming, the present results may be seen as an extension of the same process. It has been suggested that the observed change in I.Q. is a function of stimulation—those children in the lower stream receiving restricted stimulation compared with those in the higher stream. The routine or drill type of teaching methods, particularly in relation to problem solving also represent a restriction in stimulation in comparison with more flexible methods.

In a more recent study, Lovell[7a] included both tests of rigidity and tests of concept formation. He argued that such processes as those discussed here and by Vernon and Daniels might be indicative of normal deterioration due largely to lack of stimulation. Lovell drew his young stimulated and non-stimulated groups (comparable to present technical and modern groups) from schools in different localities where, although economic factors " were about the same as far as could be judged," definite differences in parental attitude and teacher ability existed. Just how far the results here presented and those reported by Vernon, Lovell and Daniels are a function of teaching method and how far a function of the other factors discussed is not ascertainable. Lovell emphasizes this difficulty and states that while the quality of teaching may have contributed to the differences between the groups " it is thought that the general background, attitude to life and work and the general adaptability of the two groups differed considerably "([7a], p. 207).

If this were precisely the case in the present study, it could be implied that the boys showing the greatest tendencies to rigidity may have been selected into the modern stream. However, the school practice of facilitating the transfer to the technical stream of those boys in the modern section who, together with their parents show sufficiently keen interest to request it and sufficient intention of completing the technical course, suggests that initially this is not so. On the other hand, teachers in the school are agreed that the boys in the modern section do not have the same enthusiasm for school and preparation for future employment as have the boys in the technical section.

The writer contends that in the present study with differences in teaching ability, school morale (climate) and school locality largely eliminated, and subjects matched for intelligence and age, the results suggest that teaching methods are of major importance. Even so, he is in agreement with Lovell that further research is necessary to ascertain the relative part played by the various factors in the development of rigidity and deterioration. In this regard two types of study may prove fruitful—the one relational, in which background and attitudinal factors are matched as well as as age, intelligence, teacher ability and school; the other experimental, in which matched groups from both stimulating and non-stimulating environments are subject to different types of teaching directed toward both increasing and decreasing rigid methods of problem solving and concept formation.

Finally, it is intended to imply by the foregoing discussion, not that background and attitudinal factors as emphasized by Lovell, and personality factors, such as insecurity or authoritarianism as stressed by Rokeach and others, have no part in the manifestation of Einstellung rigidity, but that intelligence and educational training are more salient than has hitherto been acknowledged.

* * * *

ACKNOWLEDGEMENTS.—Appreciation is expressed to the London County Council for permission to work in one of its schools and to the head teacher of the school who generously encouraged the writer in the course of the study, which was aided by a grant from the Central Research Fund of the University of London.

VI.—APPENDIX.

(i) *Problems.*

Practice Problems				
A	9	2	—	get 5
B	6	17	4	get 3
C	7	16	4	get 3
D	8	28	4	get 12
Test Problems				
E1	8	18	3	get 4
E2	20	59	4	get 31
E3	9	16	3	get 1
E4	18	43	10	get 5
C1	7	18	4	get 3
C2	23	49	3	get 20
Extn.	28	76	3	get 25
C3	14	36	8	get 6
C4	18	48	4	get 22

The problems are tabulated. The complete verbal presentation of problem E1 would be:

Given: an empty 8 quart jar, an empty 18 quart jar and an empty 3 quart jar measure 4 quarts of water.

(ii) *Instructions.*

The Einstellung problems were administered according to the instructions given in Luchins' *Monograph*, 1942, pages 2 and 10.

" Now I want you to work some problems which are probably different from ones you have done previously.

" Your task is to obtain certain quantities of water using jars of certain sizes. In obtaining the desired quantities you may use any of the jars you like but the quantities must be obtained by measuring and not guessing.

" Here is an example, watch while I demonstrate it for you.

" The first example was demonstrated—the answer being written in three ways using arrows, in figures and in words.

" The remainder of the practice and example items were completed, both methods being demonstrated for the ' critical ' example.

" Once you have started work, do not ask questions or talk. Should you need to speak to us about anything, raise your hand and one of us will come to you. In this way you will not disturb anyone else.

" Remember, the answers may be obtained by using any of the jars you like, but the answers are obtained by measuring and not by guessing."

VII.—REFERENCES.

1 AINSWORTH, L. (1953). *A Study of Rigidity.* Unpublished Ph.D. Thesis, University of London Library.

2 APPELZWEIG, D. G. (1954). Some determinants of behavioral rigidity. *J. Abnorm. Soc. Psychol.*, 49, 224-228.

3 CATTELL, R. B., and TINER, G. (1949). The varieties of structural rigidity. *J. Pers.*, 17, 321-3241.

4 CATTELL, R. B., and WINDER, A. E. (1952). Structural rigidity in relation to learning theory and clinical psychology. *Psychol. Rev.*, 23-28.

5 COWEN, E. L., and THOMPSON, G. (1951). Problem solving rigidity and personality structure. *J. Abnorm. and Soc. Psychol.*, 46, 165-176.

6 DANIELS, J. C. (1955). Streaming in junior schools. Letter in *Times Educ. Supp.*, 29th July.

6*a* LEVITT, E. E. (1956). The water jar Einstellung test as a measure of rigidity. *Psychol. Bull.*, 53, 347-370.

7 LEVITT, E., and ZEHLEN, S. L. (1953). The validity of the Einstellung test as a measure of rigidity. *J. Abnorm. and Soc. Psychol.*, 48, 573-580.

7*a* LOVELL, K. (1955). Intellectual deterioration in adolescents and young adults. *Brit. J. Psychol.*, 46, 199-210.

8 LUCHINS, A. S. (1942). Mechanization in problem solving : the effect of Einstellung. *Psychol. Monog.*, 84, No. 6 (whole No. 248).

9 LUCHINS, A. S. (1951). On recent usage of the Einstellung-effect as a test of rigidity. *J. Consult. Psychol.*, 15, 89-94.

10 ROKEACH, M. (1948). Generalized mental rigidity as a factor in ethnocentrism. *J. Abnorm. and Soc. Psychol.*, 43, 259-278.

11 VERNON, P. E. (1955). The assessment of children. *University of London Institute of Education Studies in Education* No. 7, pp. 189-215. London : Evans.

(*Manuscript received* 31*st May*, 1956.)

A NOTE ON THE USE OF THE SCHONELL AND GATES READING TESTS IN THE FIRST YEAR OF THE JUNIOR SCHOOL

BY M. L. KELLMER PRINGLE AND M. D. NEALE
(Remedial Education Centre, Institute of Education, University of Birmingham.)

SUMMARY.—Two of the Gates Reading Tests were administered to 180 children in their first year of the junior school and the results were compared with those obtained from two of Schonell's Tests. In view of a number of advantages such as easier early gradients and greater attractiveness, they are recommended as alternative or additional tests for diagnostic purposes. As the norms of the Gates Word Recognition Test are too easy for English children, an adjustment is suggested.

I.—INTRODUCTION.

TESTING reading attainment in the first year of the junior school, though often desirable for purposes of screening, or grouping within the class, is beset by difficulties where dull or educationally retarded children are concerned.* On a group test of comprehension such as Schonell's Silent Reading Test A,[7] they achieve at best only two or three items, while at worst they fail to understand even the instructions, returning blank or otherwise useless papers. Similarly, on an individual test, such as Schonell's Graded Reading Vocabulary Test, the dull or retarded young child rarely manages more than the first ten words and often refuses to attempt the others ; moreover, success on these first ten words seems to some extent to depend on the reading material and teaching methods used in the infant school. Thus, for the very children for whom it is most essential to have reliable information, the results are most meagre, partly because these tests do not call forth a sufficient degree of motivation and partly perhaps because the gradient is too steep. In view of these limitations of the Schonell Tests, it was decided to investigate the value of alternative tests measuring similar aspects of reading ability.†

As the Gates[3] Reading Tests, which are being widely used in America, are now more readily available in this country, it was planned to investigate the value of two of his Primary Reading Tests.‡ The first was the Word Recognition Test which takes fifteen minutes to administer and the second was the Paragraph Reading Test needing twenty-five minutes. There are three equivalent forms for each test. Designed for children, who, in American schools are in the second half of Grade 2 or in Grade 3, they are likely to be suitable for children aged 8-9 years in this country. The Word Recognition Test has forty-eight items, each consisting of four words and an illustration and the pupil is asked to put a circle round the word that tells most about the accompanying picture. The vocabulary becomes progressively more difficult and the group of words increasingly similar in detail and configuration. It is rare for 8-9-year-olds to complete the whole test correctly, but those who do have, according to Gates, acquired a sufficient vocabulary to read popular adult literature. The score on the test is computed by subtracting from the total number of items marked correctly, one-third of the number marked wrongly ; omitted items are disregarded. The highest possible score is 48.

---

* Experience at the Remedial Education Centre over the past five years, which was corroborated by the findings of a research project, part of which is reported in this paper, indicated the need for a new oral reading test, particularly for the young or very poor reader. To devise and standardise such a test M. D. Neale was appointed as a Research Fellow in the Education Department. Her work is now completed and the new test will be published in the near future.

† McLaren,[5] Lambert,[4] Doull,[1] Richardson,[6] Fleming,[2] and others have devised tests to overcome these difficulties, but none fulfil all the criteria for assessing the reading of young or very retarded children. McLaren, Lambert, Doull and Richardson supply norms for a limited age range only and other tests such as Fleming's " Kelvin Measurement of Reading " are not an improvement on Schonell's tests regarding their appeal to the poor reader.

‡ Obtainable from the National Foundation for Educational Research, delivery time being approximately 2 months. They are published by the Bureau of Publications, Teachers' College, Columbia University, New York, 1942.

The Paragraph Reading Test measures the ability to read sentences and paragraphs of increasing length and difficulty " with full and exact understanding of the whole."* It consists of twenty-four items, each accompanied by an illustration which the pupil has to mark according to the instructions in the paragraph he has read. The last three items are designed to give plenty of head room for the best readers. The time allowed is generous since the test measures primarily the range, accuracy and level of reading ability rather than speed. The score is the number of directions executed correctly. The last eight paragraphs, consisting of two parts, can earn two credits, thus making 32 the highest possible score.

II.—Testing procedure.

A complete age group in two different junior schools was tested with the Gates Reading Tests described above and with Schonell's Graded Reading Vocabulary and Silent Reading Test A, as well as the Moray House Picture Intelligence Test. At the time of the experiment the children were in their first year, having spent approximately 1½ terms in their new schools. The schools were chosen because they serve a wide catchment area, drawing on streets with mean back-to-back houses as well as on middle-class and professional homes. School A has a two-stream entry and the total intake from the infants' department was 102 children, of whom 90 completed the whole battery of tests. School B has a three-stream entry and during the year of the experiment 107 children had been promoted into the junior school. Of those, again 90 completed the test battery. The mean age of the children in School A was 8 years 1 month, in School B, 8 years 2 months. Both head masters were aware of and concerned with the problem of testing young juniors and were able to make arrangements for very satisfactory testing conditions. No more than 20 children took part in the group tests at any one time.

III.—Results.

(*a*) *Intelligence and Attainment Tests.*—The results of the Moray House Picture Intelligence test indicate that both groups are slightly superior to the average of the Moray House sample. School A shows consistently wider ranges of Reading Quotients, but there is close correspondence between means and standard deviations obtained in the two schools. The means of both the Gates Word Recognition and Paragraph Reading Tests are higher than those of the comparable Schonell Tests. (Table 1.) In both schools correlations are highest between the Schonell and Gates Word Recognition Tests and somewhat lower between the comprehension reading Tests. Correlations between the Intelligence Test and all the reading tests are consistently and considerably lower (Table 2).

---

* *Manual of Directions*, Gates Advanced Primary Reading Test.

TABLE 1

LEVEL OF INTELLIGENCE AND READING ATTAINMENT (N=180).

Test	Range R.Qs.	Mean Q	S.D.
Moray House Pict. Intell. Test :			
School A	73–142	104·8	13·3
School B	73–134	102·58	13·7
Schonell Grd. Read. Voc. Test :			
School A	63–166	100·52	21·62
School B	72–153	102·83	17·51
Gates Word Recog. Test :			
School A	78–149	112·58	16·14
School B	82–136	112·36	12·77
Schonell Silent Read. Test A :			
School A	79–160	103·88	19·67
School B	80–153	105·41	18·33
Gates Paragraph Read. Test :			
School A	80–163	106·13	14·76
School B	80–154	107·46	14·90

TABLE 2

CORRELATIONS BETWEEN THE SCHONELL AND GATES READING TESTS AND THE MORAY HOUSE INTELLIGENCE TEST.

	Sch. G.R.V. and Gates W.R.	Sch. S.R.A. and Gates Para. R.	Sch. G.R.V. and Moray H.	Sch. S.R.A. and Moray H.	Gates W and Moray H.	Gates Para. and Moray H.
School A N=90	·83	·74	·37	·33	·43	·39
School B N=90	·82	·76	·37	·46	·52	·54

(*b*) *Word Recognition Tests.*—In both schools a higher mean reading age was obtained on the Gates Test (Table 3). It has a lower ceiling than Schonell's Test, on which none of the children scored closely to the possible maximum while a few managed to do so on the Gates ; another indication is the bunching of scores at the top (Table 4). At the same time the Gates Test appears to be less steeply graded in the early stages. Most of the 180 children attempted to do the first two pages (containing thirty-two items) on the latter Test, while quite a number would not try any word beyond the first three lines (fifteen words) of the Schonell Test.* This may be due not only to the easier gradient at the beginning, but also to the greater attractiveness of the Gates Test which made the children more interested and willing to persevere.

---

* For example, an educationally retarded boy of average intelligence (I.Q. 95) managed only 12 words on Schonell's Test and refused to try any others. On the Gates test he scored correctly 28 words and attempted another 6.

TABLE 3

RANGE OF SCORES, MEAN SCORES AND MEAN READING AGES.

Schonell Grad. Reading Voc.			Gates Word Recog.		
range of scores	mean score	mean R.A.	range of scores	mean score	mean R.A.
1–97	31·69	8–2	1–47	28·50	8–10

TABLE 4

FREQUENCY DISTRIBUTION OF SCORES (GROUPED IN INTERVALS OF 5 POINTS) (N = 180). WORD READING.

Grouped Scores	Schonell	Gates
85—89	1	—
80—84	0	—
75—79	1	—
70—74	0	—
65—69	1	—
60—64	8	—
55—59	4	—
50—54	12	—
45—49	12	4
40—44	14	29
35—39	17	37
30—34	21	33
25—29	25	18
20—24	18	13
15—19	23	16
10—14	12	12
5—9	7	13
0—4	4	8

(*c*) *Comprehension Test.*—In both schools the scores cover almost the whole possible range (eighteen in the Schonell and thirty-two in the Gates Tests) and again the mean reading age obtained on the Gates Test is somewhat higher (Table 5).

TABLE 5

RANGE OF WORDS, MEAN SCORES AND MEAN READING AGES.

Schonell Silent Read. A			Gates Paragr. Read.		
range of scores	mean score	mean R.A.	range of scores	mean score	mean R.A.
0—18	7·45	8—3	1—31	15·50	8—7

The Gates Test has more headroom: no child succeeded in gaining the maximum score and only one child in each school gained scores just below it; in comparison two children gained maximum scores on the Schonell Test and twenty-one children scored successfully in the preceding two categories. On the other hand the Gates Test is less steeply graded in the early stages so that there is better discrimination at the bottom end. That there is less bunching of low scores can be seen from the fact that thirty children made one or no score on the Schonell Test as against only one child on the Gates (Table 6).

TABLE 6

FREQUENCY DISTRIBUTION OF SCORES (GROUPED IN INTERVALS OF 2 POINTS) (N=180). PARAGRAPH READING.

Grouped Scores	Schonell	Gates
30—31	—	2
28—29	—	2
26—27	—	4
24—25	—	16
22—23	—	11
20—21	—	21
18—19	2	21
16—17	4	21
14—15	17	13
12—13	17	14
10—11	23	17
8—9	26	9
6—7	30	15
4—5	15	7
2—3	16	6
0—1	30	1

The easier early gradient of the Gates Test enabled the poorest readers to make appreciable scores. Similarly, an analysis of scores below 10 shows that 131 children failed to achieve more than ten correct answers on the Schonell Test while the comparable number on Gates was 43 children (Table 7). Lastly, all children attempted at least the first page of the Gates Test, probably also because of its greater appeal to young children, whereas many did not attempt any item beyond the first or second on the Schonell Test.

TABLE 7

NUMBER OF CHILDREN SCORING LESS THAN 10 ON SCHONELL SILENT A AND GATES PARAGRAPH READING TEST.

	Scores	0	1	2	3	4	5	6	7	8	9	10	Total
No. of children	Schonell	13	17	9	7	4	11	13	17	13	12	15	131
	Gates	0	1	3	3	5	2	7	8	5	4	5	43

IV.—DISCUSSION AND GENERAL CONCLUSIONS.

Any conclusions drawn from this investigation must necessarily be tentative because of the smallness of the sample.

(*a*) *Word Recognition Tests.*—It seems that the Gates Word Recognition Test is considerably easier for English children than Schonell's. To some extent this may be due to the fact that American children start school a year later. While, in the long run, this may not have a disadvantageous effect on reading attainments, at this age level lower standards would be expected. Factors of attitude and motivation may also have contributed to the consistently higher scores on the Gates Test. To judge from the children's behaviour during the individual administration, there is little doubt that it is much more attractive to them than Schonell's and thus likely to call forth a greater degree of effort and co-operation.*

The Gates Word Recognition Test seems to have at least three advantages for testing children at this age : firstly, it does not require an oral performance, but lays emphasis on the understanding and recognition of words. At this stage, a child is likely to recognise in print a considerably larger number of words than he can actually read aloud with the correct pronunciation. Secondly, the pictorial layout of the test, in addition to making it more attractive and interesting, is an aid to recognition. A child may well know that a word belongs to a certain illustration without being able to read it correctly aloud. Thirdly, the vocabulary of the Gates Test appears to be easier and less steeply graded in the early stages than that of the Schonell Test so that fuller results are available for the poor reader.

Although the norms of the Gates Word Recognition Test are too high for English children, it can be recommended for use as an alternative or in addition to Schonell's, particularly for diagnostic purposes. The results give an indication of the range of words a child has learned to recognise, and this can be made a useful basis for further teaching or remedial work. Since it can also be given as a group test, it is a useful screening device for the poor reader (R.A. below 7 years) who can manage very little on Schonell's Silent Reading Tests and whose teacher is too busy to administer individually the Schonell Reading Vocabulary Test. Lastly, its greater attractiveness leads to higher motivation and thus better co-operation from poor and discouraged readers.

To make the norms more comparable it is suggested that twelve months are subtracted from reading ages obtained on the Gates Test. On our sample of 180 children, this adjustment led to the following differences : with scores unadjusted ninety-nine children obtained Reading Ages on the Gates Test twelve months or more in advance of Schonell's while the reverse was the case with five pupils. Subtracting twelve months changed these figures to eighteen children in each case.

(*b*) *Comprehension Tests.*—It would seem that the Gates Paragraph Reading Test is not materially easier than the Schonell Silent Reading Test A ; it can, therefore, be considered a suitable alternative.† Again the illustrations make the former test more attractive, thus aiding motivation. Young and backward children, whose printing or writing is often slow and laboured, grow less discouraged on Gates Test, where answers require only a cross, line or circle. Although the choice of vocabulary in a few paragraphs appears unsuitable for English children, the simple method of answering questions and the attractive layout induced many to complete the whole test within the time limit.

---

* The importance of motivation when testing reading attainment in young children was demonstrated by McLaren[5] using a similar test of her own construction. 82 per cent. of the testees aged 5 to 6 years managed to achieve some scores.

† Various adjustments to make the scores more closely comparable, did not greatly improve the position. Thus, in the total sample there were 28 children who obtained Reading Ages on the Gates Test 12 months or more in advance of Schonell's, while the reverse was the case for 22 pupils ; subtracting 3 months from the Reading Ages obtained on the Gates Test changed these figures to 9 and 29 children respectively.

VI.—REFERENCES.

1 DOULL, D. (1947). *The Construction of a Silent Reading Group Test for Children Aged 7 to 8 Years.* Unpublished Thesis for B.Ed., Edinburgh.

2 FLEMING, C. M. (1931). *A Survey of Reading Ability.* Unpublished Thesis for Ph.D., Glasgow.

3 GATES, A. I. (1947). *The Improvement of Reading: A Programme of Diagnostic and Remedial Methods.* New York: MacMillan.

4 LAMBERT, C. M. *Seven Plus Assessment (The Northumberland Series).* University of London Press.

5 MCLAREN, V. M. (1944). *A Test of Infant Reading.* Unpublished Thesis for Ed.B., Glasgow.

6 RICHARDSON, J. A. (1950). *A Factorial Analysis of Reading Ability in 10-year-old Primary School Children.* Thesis for M.A., Birmingham, summarised in *Brit. J. Educ. Psychol.*, 1950, 20, 200-1.

7 SCHONELL, F. J. and F. E. (1950). *Diagnostic and Attainment Testing.* Oliver and Boyd.

(*Manuscript received 20th October*, 1955.)

SYMPOSIUM: THE USE OF ESSAYS IN SELECTION AT 11+

V.—COMMENTS ON " ESSAY MARKING EXPERIMENTS : SHORTER AND LONGER ESSAYS "—BY D. M. EDWARDS PENFOLD

BY D. N. LAWLEY
(Lecturer in Statistics, University of Edinburgh.)
AND
A. E. G. PILLINER
(Lecturer in Experimental Education, University of Edinburgh.)

IN her article,* Mrs. Penfold makes a useful contribution to the researches concerned with the difficult problem of consistency in essay marking. She uses analysis of variance to evaluate and interpret the results of the experiments she describes. The experimental designs and the interpretations of some of the tables published in the article appear, however, to be open to objection, and there are other logical difficulties also.

Table 1 of the article presents a two-way analysis of variance of the marks assigned by each of fifteen independent examiners to short essays written by each of 165 children. Separate comparisons of the mean squares for examiners and candidates with the mean square for the (examiner × candidate) interaction, used as error, indicate that both examiner and candidate mean squares are significant. Mrs. Penfold then goes on to say, " It is of interest to observe that the ' between candidates ' variance ratio is lower than that obtained for 'between examiners,' and a simple interpretation of the comparison is that examiners vary more in their average marks than do the candidates in theirs." This interpretation we consider invalid and misleading. The value of the variance ratio (Snedecor's F) in the analysis of variance depends upon the numbers of examiners and candidates, and without taking these into account, little can be deduced from it except the straightforward verdict of significant or not significant at a chosen probability level. If one wishes to study the relative effects of examiner, candidate and error variability, one must obtain the estimates of the variance components which are independent of the numbers of examiners and candidates.

From Table 1 we obtain estimates of the components of variance as follows :

Mean square for examiners $=211\cdot29$, estimating $(165\sigma_E^2+\sigma_e^2)$.
Mean square for candidates $= 60\cdot64$, estimating $(15\sigma_C^2+\sigma_e^2)$
Mean square for error $=1\cdot33$, estimating σ_e^2
(E=examiner ; C=candidate ; e=error).

Solving, we obtain

1.273 is an estimate of σ_E^2.
$3\cdot954$ is an estimate of σ_C^2,
1.330 is an estimate of σ_e^2.

---

*PENFOLD, D. M., EDWARDS (1956) : "Essay Marking Experiments." *Brit. J. Educ. Psych.*, 26, 128-136.

To illustrate the point that the variance ratio (F) depends on the number of degrees of freedom, let us suppose that exactly the same estimates of the variance components had resulted from an experiment in which 20 candidates had written essays, each of which had been marked by 100 examiners. The mean squares (variance estimates) would have been

$$MS_E = 20 \times 1{\cdot}273 + 1{\cdot}330 = 26{\cdot}790,$$

$$MS_C = 100 \times 3{\cdot}954 + 1{\cdot}330 = 396{\cdot}730,$$

$$MS_e = 1{\cdot}330, \text{ as before.}$$

Hence $F_E = 20{\cdot}143$ (99 and 1,881 degrees of freedom),

and $F_C = 298{\cdot}293$ (19 and 1,881 degrees of freedom).

Thus, on Mrs. Penfold's argument, the candidates would have appeared much more variable than the examiners, although, since precisely the same variance components are concerned, they are, in fact, no more variable than before. It is clear that simple inspection of the F's without regard to the numbers of the degree of freedom concerned gives no indication of the relative variabilities of examiners and candidates.

In assessing the efficiency of the marking procedure it is more profitable to study the relative magnitudes of σ_C^2, σ_E^2 and σ_e^2. On statistical grounds the aim should be to reduce σ_E^2 (difference in marking standards) and σ_e^2 (disagreement over order) as far as possible. It should be observed however that considerations other than statistical are relevant here. Concentration on the more readily measurable aspects of essays may result in close agreement at the expense of other and perhaps more important aspects.

The interpretations put upon Tables 3 and 4 are also open to objection. Variance ratios (F's) are compared both within Table 3 and between Tables 3 and 4, which, as we have shown, is not a valid procedure. Further, Table 3 reports the analysis of the marking of eight scripts written on one subject; eight scripts written on a second subject; and nine scripts written on a third, all twenty-five candidates being different and all scripts being marked by fifteen examiners. The inequality in the number of essays written on each subject is stated to result in some degree of approximation. This is not the case; the design is orthogonal and when written out in full is of the following type:

		d.f.	
(1)	Between Subjects	2	
(2)	Between Candidates within Subjects	22	
(3)	Between Candidates		24
(4)	Between Examiners	14	
(5)	Subject × Examiner	28	
(6)	Candidate × Examiner within Subjects (Error)	308	
(7)	Within Candidates		350
(8)	Total	374	

In Table 3, rows (5) and (6) have been pooled to give an 'error' mean square of 4·50 with 336 degrees of freedom. The mean square for Subjects (103·1) is compared in the table with this pooled error mean square (4·50) and the resulting value of F (22·9) is taken as indicating significance of the differences among Subject means. In fact, the comparison is invalid. The Subject mean square can only be compared with that for Candidates within Subjects, since the subjects are written on by different groups of candidates. It is doubtful, however, if even this comparison is valid. We are told that the twenty-five scripts taken were " judged by the chief examiner to cover the whole mark range." The comparison is legitimate only if the groups were chosen at random, and it seems that this was not the case. Actually, if the groups are treated as though they *were* selected at random, no significant difference will be found, since the value of F is less than 1. The same objection applies to Table 4 also.

(*Manuscript received* 11*th October,* 1956.)

REPLY TO COMMENTS ON "ESSAY MARKING EXPERIMENTS"

By D. M. EDWARDS PENFOLD
(Institute of Education, University of London).

I SHOULD begin by reassuring readers that the interpretation of a sequence of analysis of variance tests is not subjective. Statisticians do not, in fact, pay the figures extra and make them mean what they like ! I think Drs. Lawley and Pilliner have misunderstood the wording on which they base their first criticism, as they have taken a remark about relative variabilities out of a context which went on "*However*, the probabilities corresponding to the variance ratios F, etc. . . . " ; and together these were intended to imply that F ratios were not to be compared at their face values but against a fixed frame of reference of probabilities determining their relative significance on the assumption of the truth of the null hypothesis, $\sigma_E^2=0$, $\sigma_C^2=0$. Unfortunately, the Scotch mist which I apparently produced seems to have developed into a London pea-souper. If my critics had in mind to demonstrate that values of F for fixed levels of probability are different for various pairs of values of degrees of freedom, they have only to refer to published tables of F, based on the above null hypothesis. Instead, they appear to have pursued a notion of an alternative hypothesis, in which σ_E^2 and σ_C^2 have non-zero values, obtained by expressing the expected values of the variances as $165\sigma_E^2+\sigma_e^2$, etc. Clearly, these are estimates, a practice against which we are warned by such authors as Brownlee[1] and Snedecor,[2] who favour the method of expression in terms of variance components on the grounds that it clarifies the sequence of tests, but point out that sampling errors may be high in the estimates of variance. The use of such *estimates* (to three decimal places) as algebraic *constants* to obtain new F values (expressed to six significant figures) debars us from using fixed probability levels as the common basis on which we might consider comparing them with the old ones. One assumes that it is more likely that they have in mind the setting up of a new hypothesis in which these are the *population parameters*, and that they are comparing the *expectation* values of two sets of F's, no longer unity as in the original null hypothesis. In this event they are wrong to attach numbers of degrees of freedom to such values, as they are no longer sample estimates.

In regard to the second point, the introduction of a comparison between subjects of essays was quite incidental, and it should be noticed that I carefully refrained from commenting on the *significance* of the variance ratios below the tables 3 and 4, to which the writers take exception, or in the final conclusions. The only consideration at the time was to make the sample of essays representative ; but in view of the fact that it was endeavoured to include a full range of ability in each topic, the design was probably nearest to one in which three *matched* groups of nine candidates wrote essays on different topics. The breakdown suggested in the criticism would not be appropriate to this situation ; neither did I feel that the data justified a full three-way classification treatment using, perhaps, the " missing plot " techniques of agricultural statisticians to estimate the missing values. I have, therefore, called the sum of squares between subjects an approximation, but I believe the variance estimate to which it leads may at least be used as an indication of disparities between subject performance.

---

[1] BROWNLEE, K. A. (1948). *Industrial Experimentation.* H.M. Stat. Office, p. 120.
[2] SNEDECOR, G. W. (1946). *Statistical Methods.* Iowa State College Press, p. 230.

The subsequent articles by Peel and Armstrong and Wiseman in the series, published in the issue November, 1956, present validation studies, and both claim some significant results. It is, of course, evident that if the essay mark, with all its inconsistencies, may be shown to be a generally good predictor of academic success, results such as mine would lose some of their force. However, Peel's article confines itself to a criterion based on assessments of English and Modern Languages only, and, though the importance of these subjects is not denied, one feels that prediction of other parts of the curriculum might be equally if not more increased by the introduction of some other test or task in the selection procedure, and clearly the latter is not capable of indefinite extension.

In this study the essays were marked only once, but any errors in the marks on this account would probably reduce the precision of the experiment, so that it is likely that those schools showing a significant gain in the prediction of the Modern Language and English criterion would maintain this position under more restrictive conditions of marking. In three out of eleven schools, however, there was no significant gain, and in others the gain was significant at the 5 per cent. level only. The cause of such variations between schools may be bound up with their curricular variations, or the nature of the assessments on which the prediction is based; at worst, less new work may have been done in those schools showing the essay as a good predictor. Since I put in a plea for more attention to be given to the inferential aspects of numerical data, I felt that the evidence presented in Table 2 was inconclusive. I calculate, however, that there is a chance of about one in eight that there would be six or more schools where the maximum prediction would be by the MHI-EC battery, in a situation where differences between one pair and another were due to small sampling errors only rather than to any systematic effects. This is not as low as the conventional level of significance, though it may be suggestive.

Dr. Wiseman's article challenges my remarks on the question of administration, and suggests that it may only be necessary to mark the essays of borderline cases. The subsequent four markings of the essays could not, it appears, be begun until the borderline cases were settled. Are there education authorities which are so patient? I hope that the author will not take it amiss if I mention that his remark on page 172 to the effect that we must show that "the inclusion of the essay does not *lower* the predictive power of the essay," followed by the comment on the numerical evidence on page 178 that the multiple correlation is increased by the inclusion of the essay in five cases of out of six, might be a little misleading to those whose statistics is less competent than his own. The multiple correlation will, of course, *never* be lowered by the inclusion of an additional variable, and it would probably be raised slightly (if not significantly) if one added to the predicting variables "the number of cups of tea drunk by the parents during the week preceding the examination."

* * * *

ACKNOWLEDGEMENTS.—My thanks are due to the Editor for allowing me a little extra space in which to make these remarks.

BOOK REVIEWS

H. S. BROUDY and E. L. FREEL (1956). *Psychology for General Education.* New York, London, Toronto : Longmans, Green and Co., pp. 456, 35s.

J. B. STROUD (1956). *Psychology in Education.* New York, London, Toronto : Longmans, Green and Co., pp. 617, 35s.

H. C. LINDGREN (1956). *Educational Psychology in the Classroom.* New York : John Wiley and Sons, Inc. ; London : Chapman and Hall, Ltd., pp. 521, 40s.

As their titles indicate, these three American books are concerned with psychology in relation to education. The title of the first, however, might possibly mislead. Its authors' contention is that psychology should be studied by every educated person ; and the book is intended to be a contribution towards, rather than a study of, this liberal education. It is not a work on ' educational psychology ' in the customary sense of that term. For example, in accordance with their purpose the authors have omitted " detailed treatment of topics that bulk large in educational psychology and child development courses." (There are, nevertheless, two chapters on Learning, which give a useful if brief account of the major theories.) It reads well, but one senses the lack of that peculiar precision and punch that can be given to a book that is addressed to a specific body of readers—such as teachers-in-training. The same lack makes difficult any assessment of the book in respect of the relevance of its material, for the " beginning student in psychology " scarcely provides adequate criteria. Lecturers in education departments and training colleges probably would prefer to direct their students to publications dealing specifically with the psychology of education, or to refer them to original sources relevant to their particular interests. It might serve as a useful introduction for students reading for a degree in psychology, an introduction to the breadth rather than the depth of the subject, though many tutors probably would prefer that that function be subserved by a lecture course rather than by a book.

In contrast, Stroud's book is very much concerned with the psychology of education, though from a particular point of view. It " may be said to be written from a sociological standpoint and to represent an approach to a social psychology of education." In the chapter on Reading, for instance, there are sections on the history of reading, reading readiness, eye-movements, and the improvement of adult reading ability, but no detailed discussion of what are usually known as ' teaching methods ' (phonic, look and say, etc.). It is in effect a background book, providing a context for ' educational psychology ' in the narrow sense of that term. In the reviewer's opinion, this view is to be welcomed ; it is as true of psychology as of any other subject that one may be knowledgeable about it without being educated in it. The book is, nevertheless, an advanced one, not because the material of it is difficult in itself, but rather because it presupposes a fair knowledge of the ' narrower' educational psychology to which it is a background. This 1956 edition is a revision of the 1945 edition ; much has been re-written and four new chapters added—including one entitled " The Profession of Teaching." It may be recommended to students reading for a degree in education. Its size might preclude its recommendation for training college courses, but tutors would find it both a useful source and a stimulating guide.

Educational Psychology in the Classroom, so far as many readers of this *Journal* are concerned, is likely to have the greatest appeal of the three books under review. It constitues a genuine and realistic attempt to relate psychological principles to the everyday work of the school, and on the whole the attempt succeeds. There is neither an over-simplification of psychological principles nor a specious invention of improbable situations to illustrate their applicability. On the contrary, it is lavishly illustrated with photographs, anecdotes, case-studies, and even cartoons that are highly relevant—and some of which are more than brilliant. (One remembers the cartoon of the tough junior who has been made to write on the blackboard, " I will not be maladjusted.") The book is convincing. It covers a wide field, from the more formal learning situations to emotional health, from the evaluation of learning to the psychology of being a teacher, from current psychological concepts of learning to a discussion of the traditional theories of learning that are implicitly held by many people. One would like to recommend this book to all students of education, but two features, shared by a great many American publications, make one cautious. First, its length ; since it is unlikely that the British lecturer or tutor would wish to recommend or prescribe only one book, even in psychology alone, can we reasonably expect students (especially those in training colleges) to read many books of this length ? And if we make one long book the corner-stone of the reading programme, should it be one that is conceived and illustrated in a foreign idiom ? That is the second point. For all the undoubted quality of Lindgren's book, one is forced to ask to what extent applied psychology is exportable, particularly in the field of education. Many of the examples themselves have to be translated out of an idiom which, to the average British student, must sound somewhat unreal. Furthermore, the present extent of educational psychology makes one doubt the expediency of trying to wrap it all up in one book. These doubts, of course, apply to a great many books. With these tentative provisos, Lindgren's book can be warmly recommended to all who are concerned with the psychology of education. J. M. THYNE.

W. M. CRUIKSHANK (edit.) (1956). *Psychology of Exceptional Youth.* Staples Press, pp. xiii+594, 50s.

Though Cruikshank starts by announcing that this symposium " is devoted solely to the psychological consideration of the influence of physical deviation upon the normative growth and development of young people," in the next sentence he defines " exceptional child " as " one who, by reason of a physical *or intellectual* deviation, is considered unique among children." The book conforms to the second definition. It has chapters on the Somatopsychology of Physical Disability, Psychological Assessment, and on all the major types of handicapped children. There is a good concluding chapter on psychotherapy in this field, but the isolated chapter on the Psychology of Gifted Children, acceptable though it is, seems out of keeping with the general theme.

Symposia are frequently uneven in quality, but in this case a high standard is maintained in chapter after chapter. Most of the contributions are balanced surveys of their topic, being particularly good in their analysis and appraisal of research evidence ; there are often suggestions for further research. The styles—which seem remarkably alike—are much more readable and less verbose than that of some American works which find their way here. The content of the successive chapters leads to a certain amount of repetition in the psychological analysis of the situation, but this jars only occasionally.

There is a good deal of emphasis on the necessity, when testing handicapped and institutionalised children by means of tests (or adaptations) used for normal children, of realising the dangers of lack of ' commonality of acculturation .' In the last chapter there are welcome warnings to psychotherapists who engage in research to endeavour to establish control groups and to strive for objectivity of criteria. Elsewhere our attention is again drawn to the scarcity of standardised tests for detecting critical and creative abilities ; their scarcity is understandable, though regrettable.

It would be invidious to single out any chapter for special praise ; much depends on the reader's own specialisation and personal taste. This is a valuable book and well worth its high price. It is a book to be read with interest and kept on one's shelf for reference to research findings. R. R. DALE.

J. P. GUILFORD (1956). *Fundamental Statistics in Psychology and Education.* New York and London: McGraw-Hill (third edit.), pp. xi+565, 47s.

This is probably one of the most comprehensive and useful statistical reference books for the educational and psychological research worker. Professor Guilford writes very clearly, and expounds all the main statistical techniques together with the conditions under which they are applicable, in a manner which can readily be followed by any student who has grasped the elements of the subject. But he probably covers too much ground for the book to be suitable for most courses in British university departments. A short list of some of the less usual topics that he includes will give a good idea of its scope: Significance of differences between matched samples, correction of correlations for selectivity, Taylor-Russell and other methods of evaluating effectiveness of selection procedures, Doolittle technique for finding multiple correlations, discriminant functions, tests of homogeneity of variance, and some of the new non-parametric techniques for testing significance. On the other hand, the techniques of factor analysis are wholly omitted, and the chapter on analysis of variance does not extend to covariance. Nor are item analysis, test construction or other applications of statistics considered. In other words, as the title states, the book aims to provide the fundamentals. There are plenty of exercises—with the right answers—at the end of each chapter, and most of the tables that the research worker needs are given in Appendices. The typography of the statistical symbols and formulae is particularly clear.

R. J. HAVIGHURST, E. STIVERS and R. DEHAAN (1955). *A Survey of the Education of Gifted Children.* University of Chicago Press, pp. 114, 11s. 6d.

The authors give some thirty pages of sound, but rather elementary, comment on the education of the gifted and then list the 'enrichment' programmes followed by selected children in some forty American schools. Though educational psychologists in this country will learn little from the book, it may well perform a useful service in U.S.A. itself by drawing attention to the need for better provision for the education of the gifted. Here and there, however, are pages giving good advice which, though well-known to our educational theorists, is inadequately practised in many of our grammar schools, e.g., p. 20, where the authors indicate the need for emphasis on the creative and experimental, on the skills of investigation and learning, on independent work stressing originality and initiative and on first hand experiences, etc. There is a useful bibliography. R. R. DALE.

INTERNATIONAL ASSOCIATION OF APPLIED PSYCHOLOGY (1955). *Proceedings of the Twelfth Congress, London, 18th-23rd July.* pp. 143, no price given.

The International Association of Applied Psychology has grown out of the Association de Psychotechnique—a body which was primarily interested in the various branches of industrial psychology. Now educational and clinical psychology are recognised as applied fields, and were well represented at the highly successful congress held in London in the Summer of 1955. Of particular interest to educational and child psychologists are the abstracts of symposia on: " Validity of Projective Techniques," " Social Psychology of the Classroom," " The Handling of Exceptional Children," " Adolescent Delinquency," " Technical Education and Training," and " The Classification of Occupations for Vocational Guidance Purposes." Some of the individual contributions are given in full, including the addresses by Dr. C. B. Frisby (the President), Sir Frederic Bartlett, Professor Viteles and Professor Mira.

O. A. OESER (edit.) (1955). *Teacher, Pupil and Task: Elements of Social Psychology Applied to Education—A Practical Manual for Teachers.* Tavistock Publications, Ltd., pp. 180+xvi.

This book consists of reports on different aspects of the classroom situation and teaching compiled by members of the Department of Psychology, University of Melbourne, Teachers' College staffs and psychologists and research workers of the

Education Department, Victoria, under the editorship of Professor Oeser. All could be regarded as experienced teachers. It is based upon the discussions of 250 teachers attending a Refresher School. The authors were leaders of the discussion groups at the School.

It deals with the practical applications of some principles of social psychology to classroom teaching with particular emphasis on motivation, classroom organisation and method, and the conditions for efficient learning and growth. It is intended to publish another dealing with the theoretical background. The authors intend to present a practical guide to teachers to indicate the significance of the 'social climate,' of the classroom and how it may be used for better teaching. The first four chapters deal with the social aspects of teaching such as the social roles of pupil and teacher, the position of the teacher in relation to the head master, the pupil and the curriculum, the social conditions for successful learning and the classroom as a social group. Two chapters are devoted to sociometry whereby the patterns of relationships that exist in a class are revealed, and suggestions are given of what may be done by the teacher to use or change these patterns to facilitate teaching. The last four chapters have little to do with social aspects as such, but are included presumably to illustrate that the social approach is not aimless sentiment but is firmly tied to testing, problem children, examinations and the adjustment of educational and vocational problems.

There is a trend in educational thought to get away from the exclusive emphasis on individual characteristics such as intelligence, motivation, emotion, and the like, and to pay more attention to the social forces of the class with which those characteristics interest and which have considerable significance to the child's learning and all-round development. This book is representative of that trend and it also departs in some degree from current views on the child orientated curriculum and method. The authors are obviously conscious of the need to maintain standards in achievement, and they point out how these may be obtained more harmoniously and effectively by taking advantage of the social forces within the group. Here one would have referred more documentation and evidence to be given rather than more or less isolated instances of the success to be achieved in the immediate classroom situation and of the long-term effects that are supposed to follow.

The book achieves its purpose of presenting to teachers a practical guide to some principles that are highly relevant in our generally autocratic schools, which allege democratic ideals but are more remarkable for their disregard of those ideals in practice. Teachers should find it suggestive and a source of ideas for liberalising and humanising their practices. All will find it quite readable and free from the technical terms of the social psychologists. G. T. SMITH.

M. OSTERRIETH, J. PIAGET, *et al.* (1956). *Le Probleme de Stades en Psychologie de l'Enfant.* Symposium de l'Association psychologique scientifique de langue française. Genève : Presses Universitaires de France.

The purpose of this symposium is to inquire whether stages of development are artificial divisions invented for convenience, or whether they correspond to natural discontinuities in psychological growth. Articles contributed relate to stages in physiology, personality and emotional and intellectual development ; subsequent discussion is reported.

M. Osterrieth has compared eighteen different systems of stages, European and American, and finds practically no agreement in chronological periods except in the intellectual field. He attempts to reconcile the systems of intellectual stages and arrives at five major periods of development from birth to adolescence.

M. Piaget doubts the existence of general stages which include all physiological and mental functions. In his article concerning stages in intellectual development he emphasizes that stages appear in an invariable order, though their appearance may be delayed or accelerated by the child's experience or by socio-economic factors. He distinguishes the final structure characteristic of a stage from the formative processes which successively differentiate it and he introduces the idea of displacement which modifies the conception of stages. Displacement is horizontal if, within a major period of development, a child is capable of an operation in some fields but not in others until a year or two later ; it is vertical when a child learns to deal with an operation at a level of thought characteristic of a more advanced stage. M. Piaget defines three major periods of intellectual development and describes their sub-states. In discussion he stresses the need for longitudinal studies of individuals.

Researches by Mlle. Inhelder and M. Rey concerning aspects of the relation of perception and thought are reported in Part 2 of this volume. R. BEARD.

L. T. WILKINS (1956). *The Adolescent in Britain.* Central Office of Information, 25s.

This publication, being full of unvarnished facts, will be of value to research workers in any of the social sciences and to educational psychologists. The facts have been collected by a team of interviewers from a random sample of 1,390 male and 451 female adolescents aged 15-19 years. The concurrence of its findings with information from other sources shows that the sample has been accurately drawn. The facts are presented with little comment and no speculation, and the work could, with profit, be consulted by all who tend to begin sentences with the words " Young people nowadays . . ." Norms are provided under the headings of home background, education, employment, and leisure pursuits. Most of the information has been cross tabulated. 123 tables are provided, and no doubt other cross tabulations could be given in answer to specific enquiries from research workers. Those who are studying particular groups, such as maladjusted or delinquent children, will be particularly interested to know what is the ' expected ' percentage of broken homes, working mothers, over-crowding, etc., in this representative sample. Teachers will find significance in the relationship between happiness at school and stability in employment, and most readers will find that the facts about cinema attendance offer an interest basis for speculation. It is unfortunate, that in a survey of this kind there has to be so much delay between the collection and publication of the information. This survey was made in 1950, and already some of its information, for example, that about television as a leisure pursuit, is out of date. This inevitable limitation is, of course, beyond the control of the author, who is to be congratulated on a laborious task, painstakingly performed. M. D. WILSON.

PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1956). *Hearing Defects of School Childdren.* London : University of London Press, pp. x+112, 10s. 6d.

J. S. BRUNER, J. J. GOODNOW and G. A. AUSTIN (1956). *A Study of Thinking.* New York : John Wiley ; London : Chapman and Hall, pp. xi+330, 44s.

E. M. LIGON (1956). *Dimensions of Character.* New York : Macmillan, pp. xxix+497 45s. 6d.

Our Smallest Servants : *The Story of Fermentation* (1956). Folkestone, Kent : Chas. Pfizer, Ltd., pp. 33. Free to schools or libraries.

WORLD HEALTH ORGANIZATION (1956). *Expert Committee on Psychiatric Nursing, First Report.* Geneva : W.H.O., Technical Report Series, No. 105, pp. 43, 1s. 9d.

G. S. WELSH and W. G. DAHLSTROM (edit.) (1956). *Basic Readings on the M.M.P.I. in Psychology and Medicine.* Minneapolis : University of Minnesota Press ; London : Oxford University Press, 1956, pp. xvii+656, 70s.

N. O'CONNOR and J. TIZARD (1956). *The Social Problem of Mental Deficiency.* London : Pergamon Press, pp. viii+182, 30s.

R. BAKEWELL and D. FLETCHER (1956). *Growing and Reading: A Photographic Reading Scheme for the Infant School.* London: Macmillan, Stage 1 Reader, 3s. 3d.; Stage 2, 5 books, 1s. 4d. each; 12 supplementary books, 10d. each.

G. VARET (1956). *Manuel de Bibliographie Philosophique, II. Les Secinces Philosohpiques.* Paris: Presses Universitaires de France, pp. 499-1058, Fr. 1960.

INTERNATIONAL ASSOCIATION OF APPLIED PSYCHOLOGY (1955). *Proceedings of the XIIth Congress, London, 18th to 23rd July.* pp. 143. No price given.

H. TAYLOR, *et al* (1956). *A Graduate Program in an Undergraduate College: The Sarah Lawrence Experience.* Sarah Lawrence College Publications No. 6. Middletown, Conn.: Wesleyan University Press, pp. xx+119, $1.50.

B. R. BUGELSKI (1956). *The Psychology of Learning.* New York: Henry Holt; London: Methuen, pp. xiii+523, 50s.

F. CAMPBELL (1956). *Eleven-Plus and All That.* London: Watts, pp. xviii+193, 15s.

J. C. DANIELS and H. DIACK (1956). *Progress in Reading.* University of Nottingham; Institute of Education, 1956, pp. 71. No price given.

ERRATUM.

The sentence beginning on line 25 on page 9 of the article *Recent Developments in Piaget's Work* by D. E. Berlyne in the February number of this Journal, should read:

> "But 'p implies q' has both an inverse ('p does not imply q') and a reciprocal ('q implies p').

VOLUME XXVII. NOVEMBER 1957 PART 3

CONTRIBUTIONS TO INTELLIGENCE TESTING AND THE THEORY OF INTELLIGENCE

I.—ON DEFINING INTELLIGENCE.

BY T. R. MILES
(*University College of North Wales*).

I.—*Introduction.* II.—*Different senses of the word ' definition.'* III.—*Application to the study of intelligence.* IV.—*Wechsler's definition.* V.—*Burt's definition.* VI.—*References.*

> " We have first raised a dust and then complain that we cannot see."
> —BISHOP BERKELEY.

SUMMARY.—In offering what purport to be definitions of intelligence, psychologists do not always seem to have worked out what sense of the word ' definition ' they have in mind. Six possible senses of the word ' definition ' are here distinguished. Each sense is then discussed with special reference to the problem of defining intelligence. In the light of the distinctions made, the definitions of intelligence offered by Wechsler and Burt are critically examined from the point of view of methodology.

I.—INTRODUCTION.

IT is commonly thought to be a great scandal that psychologists cannot agree on a definition of intelligence. People then draw the conclusion that intelligence must be something very obscure and elusive to provoke such controversy. I do not dispute that there are many disagreements on matters of fundamental principle, but it seems to me that the issue has often been confused by *unnecessary* disputation and by argument at cross purposes.

I shall not, in this paper, offer any definition of intelligence of my own, nor shall I take sides on the question of whether a particular definition is a good or bad one. My task is the preliminary one of clearing the ground. The question towards which I wish to focus attention is : *By what arguments do we establish that one definition of intelligence is better than another ?*

I shall suggest in answer to this question that the word ' definition ' is ambiguous, and that different arguments are appropriate according to the sense in which the word ' definition ' is being used.*

II.—DIFFERENT SENSES OF THE WORD ' DEFINITION.'

In distinguishing different senses of the word ' definition ' I have relied largely on the work of Robinson (1950).† A distinction needs to be drawn in the first place between *nominal* and *real* definition.* Nominal definition is concerned in the main with the meaning of words rather than with the things for which the words appear to stand.† Robinson sub-divides nominal definition into two classes, (1) lexical definition and (2) stipulative definition. A lexical definition gives an account of how a word has in fact been used by a particular group of people ; a stipulative definition states how the speaker proposes to use the word, irrespective of how that word has been used in the past.

---

* I have used throughout the phrase " different *senses of the word* ' definition ' " in preference to the more familiar " different *kinds* of definition." Neither phrase is wholly satisfactory. The important point is that ' defining ' is not the name of a single procedure, but refers to a group of procedures having a certain " family-resemblance " (to use Wittgenstein's phrase) between them.

† This must not be taken as a suggestion that the work of earlier writers on the subject of definition can simply be dismissed. Many of the traditional ' rules ', e.g., that a definition should be *per proximum genus et differentiam specificam* seem to me not so much wrong as in need of reformulation.

In contrast, ' real definition ' is commonly taken to be definition of *things*. A real definition is supposed to tell us the ' nature of the thing defined.' It is here that some of the biggest pitfalls in argument occur. Robinson distinguishes no less than twelve different activities, all of which have been bunched, very confusedly, under the general title ' real definition.' Of these twelve I shall mention three : (1) *The search for the essence or essential nature of a thing.* This notion Robinson regards as misleading, on the grounds that there are no such things as essences in the sense given to the word ' essence ' by Aristotle. In Robinson's view, " Is it part of the essence of a swan to be white ? " is a disguised request for a nominal, not a real definition, and means no more than " If I were to see a creature otherwise like a swan but black, should I continue to give it the label ' swan ' ? " (2) " *Description plus naming.*" Robinson writes " Many so-called ' real definitions ' of the form x is yz are equivalent to the statement that : ' The character yz occurs and I call this character (or it is commmonly called) by the name ' x '." Robinson regards this as a legitimate and useful activity, but suggests that ' real definition ' is a misleading name for this activity. It suggests the hopeless search for real essences, and invites confusion with other activities also grouped under the heading of ' real definition.' (3) *The search for a key.* A definition of x, on this showing, involves a single short sentence from which follow all the things which we need to know about x. The stock example is geometry, where all the important things we know about triangles—so it was supposed—follow from the definition of a triangle as a plane figure bounded by three straight lines. This account of a triangle thus provides the key to understanding a wide range of other true sentences about triangles. Again Robinson is hesitant to call such procedure ' real definition,' partly, once more, because it might be confused with his other eleven possible senses of ' real definition,' and also because it tends to conceal from us the fact that in some cases no such key may be discoverable.

There is a further procedure, not discussed in any detail by Robinson, but playing quite a large part in modern psychology—the so-called ' operational ' definition. This is an attempt to define the meaning of a word in terms of the observations, or scientific ' operations,' necessary if that word is to form part of a true sentence. Thus it might be said that the word ' length ' requires to be defined in terms of the operations involved in measuring length, and that " This rod is 6-in. long " is meaningful in virtue of the possibility of specifying in detail the appropriate operations.‡

This classification of definitions must not be regarded as exhaustive, nor need the different activities which I have distinguished be regarded as mutually exclusive. The important point which I wish to stress is that, until we know the sense in which the word ' definition ' is being used, attempts to assess the merits or de-merits of a defintion of intelligence are liable to lead to argument at cross-purposes.

---

* Compare Burt (1947, p. 129).

† This statement requires qualification, but is accurate enough for present purposes.

‡ For further discussion, see Bridgman (1927), esp. Chapter 1.

III.—APPLICATION TO THE STUDY OF INTELLIGENCE.

(i) *The ' real essence ' of intelligence.* If we agree with Robinson that the notion of ' real essences ' is a mistaken one,* then sentences which refer to " the real essence of intelligence " are either illegitimate or in need of reformulation. Few, if any, writers at the present time speak in as many words of the ' essence ' of intelligence.† But there are plenty of people who ask about its ' real nature ' or ' real meaning,' which comes to much the same thing. The presupposition underlying such questions is that there is one thing and one only which intelligence *is*, and that the task of psychologists is to discover it. Any sentence starting " intelligence is . . . " justifiably arouses one's suspicions.

The assumption behind the mistake seems to be that every word has one settled and precise meaning, or, more strictly, that classifications of things in nature are somehow done *for* us. The truth is surely that *we* must classify as suits our purposes. For many purposes it is helpful to classify behaviour into ' intelligent ' and ' unintelligent,' but it does not follow that there is to be found in the universe one permanently existing ' thing ' which intelligence *is*.

It is important in this connection to pay attention to inverted commas. One function of inverted commas—there are others—is to indicate that a word is being mentioned as opposed to used. Thus, when I say " This table is brown," I am *using* the word ' table ' not mentioning it, and no inverted commas are needed. On the other hand, if I say " ' Table ' is the English equivalent of the Latin ' mensa '," I am *mentioning* the words ' table ' and ' mensa ' and both require inverted commas. In exactly the same way, if we are *using* the word ' intelligence,' as in " Intelligence increases up to age 15," there are no inverted commas ; but if we say, " I recommend that the word ' intelligence ' be defined in a particular way," we are mentioning the word ' intelligence,' not using it, and the inverted commas are indispensable.

When Piaget (1950, p. 7), says, " Intelligence is thus only a generic term to indicate the superior forms of organisation or equilibrium of cognitive structurings," it seems fair to point out that the word ' intelligence ' is being *mentioned* here, not *used*, and that inverted commas round the word ' intelligence ' would make his formidable statement, if not crystal-clear, at least easier. In general, it may be said that the surest way of avoiding muddle about ' real essences ' is to pay strict attention to inverted commas.

(ii) *Lexical definitions and the appeal to ordinary usage.* I propose to argue in this section that, as far as lexical definitions of intelligence are concerned, there need be no serious disputes among psychologists. It is implicit in the notion of a lexical definition that its merits should be decidable by an appeal to ordinary usage. This appeal presents certain problems which require discussion ; but even if two psychologists did in fact disagree on what constituted ordinary usage, no important theoretical consequences would follow.

I shall begin by considering certain points about the way in which the word ' intelligence ' normally functions, and I shall then attempt to remove some of the confusions and difficulties that are liable to arise in connection with discussions about the lexical definition of intelligence.

(*a*) The word ' intelligence ' is the noun of the adjective ' intelligent.'

---

* For the arguments which he gives in support of this view, see pp. 153-6. The relevant passage in Aristotle is *Metaphysics*, Z, 4-6.

† The word ' essence ' does occur, however, in the earlier literature on the subject of intelligence. Thus, Spearman (1927, p. 15) quotes an earlier writer, Bobertag, as saying, " The knowledge of the essence of intelligence is *naturally* a thing that merits profound research " (my italics).

This point seems obvious, but is none-the-less informative. If we are not careful, we are liable to suppose that all nouns refer to ' things,' whereas it is by no means obvious that the word ' intelligence ' can correctly be said to refer to a ' thing.' By using the adjective ' intelligent ' any temptation towards misleading hypostatisation (i.e., treating words as ' thing '-words when they are not) can be avoided. Instead of the unsatisfactory " What is intelligence ? " we can now ask " What is the meaning of the word ' intelligent ' ? " or, perhaps better, " How does one test if a person is intelligent ? i.e., what constitute samples of intelligent behaviour ? "

(*b*) The word ' intelligent ' may be labelled a *disposition-word*.* Here are further examples of disposition-words—' Lazy,' ' bad-tempered,' ' cheerful,' ' kind-hearted,' ' punctual.' In all these cases there is no necessary suggestion of a person's actually doing something here and now. " X is very kind-hearted " may be true even though at this moment X happens to be asleep. The suggestion is rather that a person to whom any of these adjectives applies is *disposed* to act in certain ways, i.e., that if certain conditions are fulfilled, certain behaviour will follow. Thus " X is lazy " is approximately equivalent to " If X is given any hard tasks he usually tries to shirk them." In general, sentences containing disposition-words can be replaced by sentences containing the words " if . . . then." Similarly, " X is intelligent " can be taken as equivalent to " If X is placed in particular circumstances he produces responses of a particular kind "—e.g., if he is present at a group discussion he makes appropriate remarks, if presented with a difficult crossword puzzle he can usually solve it, and so on.

I want now to introduce three further technical terms in addition to ' disposition-word.' The word ascribing the disposition, such as ' lazy ' or ' punctual,' I shall refer to as the *substrate*; the actual or possible manifestations of the disposition I shall refer to as the *exemplaries*. Thus the exemplaries of " X is kind-hearted " are the particular occasions when he is kind to people, the exemplaries of " X is lazy " are the particular occasions when he shirks tasks, and so on. Thirdly, I shall make use of the term *polymorphous*.† The concepts ' grocer ' and ' solicitor ' are, we might say, polymorphous as compared with the concept ' baker.' A baker, *qua* baker, does one thing only—he bakes. A grocer does all sorts of different things—he weighs out sugar, he sells butter, he cuts bacon, and so on. A solicitor draws up wills, advises clients, etc., etc. There is no one way to manifest being a grocer or a solicitor. Similarly, ' lazy ' and ' bad-tempered ' are polymorphous concepts as compared with ' punctual.' The exemplaries of being punctual are an unvaried series of arrivals on time, the exemplaries of being lazy or bad-tempered cover a wide range of different sorts of behaviour.

(*c*) Using this terminology we may say that ' intelligent ' is a polymorphous concept. In other words there are many different exemplaries which the substrate ' intelligent ' carries ; intelligence may manifest itself altogether differently on different occasions.

(*d*) Finally, it should be stressed that the list of exemplaries carried by the word ' intelligent ' is *open*. In other words, no one has ever made any precise legislation as to what shall or shall not count as exemplaries of the word ' intelligent ' ; nor is there any precise list laid up in heaven for the discerning to discover. Just now I mentioned two examples—making appropriate remarks in a discussion and being able to solve difficult crossword puzzles. But clearly these are only two exemplaries among many. By making stipulative definitions we can formulate a precise list of exemplaries if we wish, but 'intelligence,' as it functions in ordinary speech, carries no such list.

---

* Compare Ryle (1949), Chapter 5.

† This term, and the following examples, are due to Ryle. There is no reference to Freud's use of the word.

Now the fact that the list of exemplaries carried by the word ' intelligent ' is ' open ' may give rise to disputes in the less straightforward type of case. Sometimes, of course, disputes may be genuinely factual. Thus, if it is known that X has an extraordinary facility for rapid calculation in his head, someone may still say, " But he may not be intelligent," the suggestion being that other exemplaries carried by the word ' intelligent,' will not be satisfied. In other cases the dispute is in a sense a verbal one—a matter for linguistic decision. Spearman (1927, p. 10) quotes an amusing example in this connection : " Trabue . . . told of a woman who, although making a bad record with the tests, nevertheless became ' the housekeeper at one of the finest Fifth Avenue hotels, where she successfully directed the work of a corps of approximately fifty maids, three carpenters, two decorators, and a plumber.' He was moved to conclude as follows : ' In spite of the evidence of the tests I insist that she is intelligent '." Perhaps there is *some* factual dispute here as to how the woman would behave in different circumstances, but Trabue's words suggest that the point at issue is not how the woman is likely to behave in different circumstances, but whether ability to supervise fifty maids, etc., constitute exemplaries of the word ' intelligent.' To this question there is no uncontroversial answer, not because we are incompetent psychologists, but because the inexactness of ordinary language makes a conclusive answer impossible.*

Even, however, if a psychologist is accused in his definition of intelligence of doing extreme violence to ordinary usage, no important principle is at stake. All that he need do is to leave the word ' intelligence ' imprecise and invent a special technical term of his own. Thus, if we are sure in advance that it is informative to test a person's memory span for digits, and if, as a result, we include such an item in what we call an ' intelligence-test,' we will not be seriously worried if someone says, " But this is a test of memory, not intelligence." All that is required is to invent a new technical term and to say that *that* is what we are measuring. The important point is not whether what we measure can appropriately be labelled ' intelligence,' but whether we have discovered something worth measuring. And this is not a matter that can be settled by an appeal to what is or is not the correct use of the word ' intelligent.'

One final point should be made in connection with lexical definitions. When it is said that psychologists do not agree on the definitions of intelligence, the position may seem all the more absurd if we unwittingly assume that they disagree about its lexical definition. Psychologists in that case are guilty of using a word without explaining its meaning and perhaps without being able to do so. It is as though someone attempted to give a talk on armadillos when neither he not his audience knew what the word ' armadillo ' meant. Clearly, explanation of any new or unfamiliar word is obligatory as soon as that word is introduced.† In the case of the word ' intelligence,' however, no one, surely, wants a definition in *this* sense. In the standard sense of ' know the meaning of,' the great majority of English-speaking adults, including psychologists, know the meaning of the word ' intelligence ' already. To put the matter somewhat more precisely, they can recognise particular pieces of behaviour as constituting exemplaries of the word ' intelligent.' No one thinks that, if the word ' intelligence ' figured in the Terman-Merrill vocabulary test, most professional psychologists would fail to secure a pass. Psychologists may disagree as to what general formula, if any, will lead to significant advances in our understanding of intelligent behaviour ; but, as far as lexical definitions are concerned, there is very little that they *could* disagree about, and certainly nothing of major importance.

---

* Spearman seems to me to have shown himself extremely sensitive to this sort of problem. Compare his well-known dictum that ' intelligence ' is " a word with so many meanings that finally it has none " (1927, p. 14).

† Compare Robinson (1950, p. 41). As Robinson in effect points out, the question whether definitions should come at the beginning or at the end of an enquiry cannot be answered unless we are told the sense in which the word ' definition ' is being used.

(iii) *Stipulative definition* ; (iv) *Description plus naming* ; and (v) *The search for a key*. These three processes will be considered under the same heading since there is considerable overlap between them.

To frame a stipulative definition involves either coining a new word or announcing that one intends to use an existing word in a special way. Stipulative definitions of intelligence clearly involve the latter. In all uses the *purpose* of the stipulation needs to be considered. The need, in this case, is for concepts which help our understanding of intelligent behaviour.

' Description plus naming ' is a helpful activity in psychology, provided what is named is something worth investigating. Similarly, to ' search for a key ' is helpful, provided that the key, when we find it, really does unlock the requisite doors. Both these activities (which can legitimately be regarded as varieties of stipulative definition) involve, in effect, the commendation of a policy. Using our earlier terminology, we may say that to make a stipulative definition is to formulate a substrate—a substrate whose exemplaries are thereby assumed to be worh investigating. Differences of opinion on policy are *serious* differences, and give rise to heated controversy. In the study of intelligence, as in any other study of personality, the crucial question is the choice of substrates. Thus, the substrates ' extraversion-introversion ' are helpful if there is a suggestive association between people's scores on extraversion-introversion tests and their other independently observed behaviour. In the same way those who study intellectual differences need to produce substrates that are *worth-while*. Disagreements over *stipulative* definitions of intelligence are far more fundamental and serious than disagreements over lexical definition.

It does not, of course, follow, because the substrate ' intelligent ' is helpful for workaday purposes, that it necessarily holds the key to any great scientific advance. It may do so ; but we have no right to assume it.*

(vi) *Operational definitions*. Those who insist on the importance of operational definitions for scientific method are in effect pointing out that a substrate has meaning only in relation to its exemplaries. The relation between exemplaries and substrate is not that of effects to an unknown cause, but of a series of occurrences to a general law under which they can be subsumed. It is pointless to assume the existence of an unknown entity lying behind or beyond the exemplaries.†

Applied to the notion of ' intelligence,' this is, in effect, to say that the word ' intelligence ' does not refer to a ' real thing ' lying behind or beyond the manifestations of intelligence. Instead of " *What is* intelligence ? " or even " What does the word ' intelligence ' *really mean* ? " we need to ask instead, " How do you test—or what operations are involved in testing—whether a person is intelligent ? "

---

* Compare Heim (1954, p. 46) : " A majority of the factorists appear agreed that a clear-cut key to these problems exists and is in their hands."

† Compare Berkeley's attack on the notion of *material substance*, passim.

This question can readily be answered. Psychologists have devised *standardised* tests. It is the items in these tests (or, more strictly, the person's behaviour in producing correct responses to these items) that are regarded as constituting the exemplaries of the word ' intelligent.' Intelligence, in other words, *is what intelligence tests measure.* This definition is a stipulative one. What is being said is, in effect, that correct responses to the test items *shall be deemed to constitute* exemplaries of the word ' intelligent.' To give a tidy list of exemplaries all that is needed is to specify what particular test we have in mind.

" Intelligence is what intelligence tests measure " does not, of course, tell us what test items are *good* ones, but this is not a ground for criticism since it does not set out to do this. Whether a list of exemplaries should be as tidy as this stipulation makes them is perhaps questionable. Some would say that a substrate should only *suggest* exemplaries, not specify them to the last detail, and that details of exemplaries can be worked out and modified in the course of future research. Despite this, however, the ' operational ' approach to the study of intelligence seems to me fundamentally sound, and a great improvement on the traditional search for " what intelligence really is."

IV.—WECHSLER'S DEFINITION.

In this section I shall examine Wechsler's introductory remarks on the subject of intelligence.* I shall argue that, whatever the merits of his policy, his formulation of that policy is unnecessarily puzzling and difficult simply through failure to distinguish the ambiguities of the word ' definition.'

Wechsler begins by discussing the dictum that intelligence is " what intelligence tests measure." He refers disparagingly to it as " this circular position " (p. 3), without, apparently, having noticed, that, if in the definiens we substitute the names of *particular* intelligence tests—the Terman-Merrill or his own, for instance, any trace of circularity disappears.†

He reinforces his argument by saying that the lay person " is entirely justified in asking, ' Do your tests really test intelligence ? ' " (p. 3). This seems to me to be putting the cart before the horse. If we agree to define the word ' intelligence' as " what intelligence tests measure," then admittedly " Do your tests really test intelligence ? " becomes pointless. But it is very perverse, surely, to argue, on the grounds that " Do your tests really test ingelligence ? " is *not* pointless, that " intelligence is what intelligence tests measure " *must therefore* be unhelpful. Why anyone should even have *wanted* to say that intelligence is what intelligence tests measure is a question that Wechsler does not seem to have seriously considered.

The purpose of this dictum, as we have seen, is to warn us against asking meaningless questions about the ' real nature ' of intelligence. Having ignored the warning notice, Wechsler not surprisingly falls head-foremost into the abyss of muddle against which he was warned. " General intelligence," he says (p. 4), " like electricity, may be regarded as a kind of energy. We do not know what the *ultimate nature* of this energy is, but as in the case of electricity, *we know*

---

* Wechsler (1944, pp. 3-4).

† A further difficulty is that Wechsler does not seem to have considered for what reasons and in what circumstances a definition should not be circular. If a child asks the meaning of the word ' armadillo ' and is told " It's a—well, it's an armadillo," we would rightly condemn such procedure as uninformative. But if we are told that the symbols bc can be substituted for the symbol a, that xyz can be substituted for bc, and that a can be substituted for xyz, this is not necessarily a futile procedure. Not only may it teach us the rules for the interchange of certain words, but, in addition, provided we can relate one of the three groups to the actual word, we will then be able to relate the other two groups to the actual world also. Circularity of this sort is not necessarily a vice.

it by the things it does " (my italics). Just what use it is to postulate such energy is unexplained, and its existence or non-existence would make no difference whatever to the value of Wechsler's tests. The concepts ' intelligence' and ' electricity ' are similar (in so far as they are similar at all) not because both refer to some unknown ' energy ' but because both require to be understood in terms of their exemplaries.

Finally, what are we to make of Wechsler's own definition ? Intelligence, he tells us, is " the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment," (p. 3). In what sense is this a definition ? Is he explaining the meaning of the word ' intelligence ' to someone who does not know it ? Is he explaining how he himself proposes to use the word in future ? Or is he offering a ' real ' definition in one or more of Robinson's twelve different senses, and, if so, in which sense ?

Despite the absence of any explicit answer to these questions, some of Wechsler's intentions can be understood from the context. Clearly he is not content merely to offer his own stipulative definition. This can be inferred from the fact that he is concerned to answer the plain man's question, " Do your tests really test intelligence ? " He must therefore be concerned, at any rate among other things, with the plain man's use of the word. If this is all, however, nothing very exciting follows *whatever* answer he gives. " Do your tests really test intelligence ? " if re-phrased with more precision appears to mean, " Do those who score highly on your tests correspond with those whom we should call ' intelligent ' in normal speech ? " But if this is how the plain man's question should be understood (and it is hard to see what other interpretation is possible) nothing very helpful is being asked. In the first place the list of exemplaries carried by the word ' intelligent ' in ordinary speech is ' open '—in other words there are many different grounds for saying of a person that he is intelligent. Only if there were a clearly specified list of exemplaries would it be possible to tell whether these exemplaries tended to be satisfied significantly more often among high test-scorers than among low test-scorers. As ordinary speech gives *no* clearly specified list of exemplaries, the question " Do your tests really test intelligence ? " allows at best for only an approximate answer. This, of course, is not due to any human failing nor to the presence of any overwhelming mystery, but arises simply as a result of the imprecision of ordinary speech—an imprecision that for everyday purposes is altogether useful. Secondly, even if it were agreed that those who scored highly on the tests were *not* those who could appropriately be labelled ' intelligent,' still nothing of importance would follow. It might be misleading to label the tests ' intelligence 'tests, but they could quite well be given another name.

It follows, if the above argument is right, that Wechsler cannot be (or perhaps should not be) interpreted as attempting a simple lexical definition of intelligence—an account of how the plain man uses the word. Indeed, I think he would be somewhat surprised if someone hearing his definition for the first time replied, " Yes, I agree ; that is just what the plain man *does* mean," or " No, you are wrong ; people have used the word differently." What Wechsler says should be regarded rather as a key for understanding something. It is as though he said, " An important key for understanding people can be obtained by determining if they have an aggregate or global capacity to think purposefully . . . etc." The conclusion to be drawn is that tests which sample such capacity are to be commended. What Wechsler's definition does in effect is to commend a policy.

Now it is at this point that the crucial decisions really begin. The psychologist has to decide—partly by appeal to evidence, partly by personal

' hunch '—whether such a policy is worth following, whether in effect the Wechsler tests (which claim to sample such ability) are good and helpful ones, whether measuring the correlations of test-scores with other personality traits will lead to further important discoveries, and so on. Decisions of this sort, however, are beyond the scope of this paper. All that I have attempted to do is to clear the ground, to leave a straight issue to be decided on its merits. Appeals to the plain man's use of the word ' intelligent ' are liable to confuse the issue, and talk of underlying ' energy ' is liable to confuse it still further. If Wechsler's definition of ' intelligence ' is thought to agree with ordinary usage, it does not follow that his policy is right ; and if it is thought *not* to agree with ordinary usage, it does not follow that his policy is wrong. I have made no attempt to criticise the way in which Wechsler *operates* his policy, but he is not alone among psychologists in being better able to operate a policy than to talk about his operations.

V.—BURT'S DEFINITION.

Unlike Wechsler, Burt (1955) shows himself extremely sensitive to the sort of problem which I have been discussing in this paper. That being so, my discussion in this section will be mainly a commentary on his article rather than a criticism of it. Even in Burt's case, however, there are certain confusions in his use of the word ' definition ' which seem to me to require to be exposed. Once again, needless to say, I shall not take sides on the more fundamental question of whether the definition is a good one.

Burt's formula is " innate general cognitive ability." What we need to examine, therefore, is the sort of argument which should be taken as relevant if this formula is to be justified.

Let us start by considering his comments on the famous 1921 Symposium. " The editor of an American journal," he writes, " submitted two searching questions about the nature of intelligence to a dozen different psychologists, and received a dozen different replies. But the varying descriptions suggested were not, as Dr. Heim and others have supposed, intended to be ' definitions ' *in the strict logical sense* " (my italics) : " they were, in the language of J. S. Mill, merely ' attempts to explain the thing ' not ' attempts to interpret the word,'." (p. 159). Unfortunately, nowhere in this article does Burt say what he means by " *the* strict logical sense " ; and this, in view of Robinson's distinction between the many different senses of the word ' definition ' is surely a defect. (What, any way, is meant by a ' logical ' sense of the word ' definition ' ? Surely the contrast cannot be with some *illogical* sense ?).

There is an even more puzzling footnote. Burt writes, " In framing his question, the editor specifically asked not how is intelligence to be defined but ' what do you conceive intelligence to be . . . ? ' etc." According to Burt, then, " How is intelligence to be defined ? " and " What do you conceive intelligence to be ? " are to be understood as different questions. This seems to me very unsatisfactory—not, indeed, because I think the questions should be regarded as identical, but because both are so thoroughly ambiguous that there is just no means of knowing if they overlap or not. Robinson's comments* seem appropriate here. " The confusedness of the concept of real definition is an effect of the vagueness of the formula ' What is x ?' For it is the vaguest of all forms of question except an inarticulate grunt. Real definition flourishes because the question-form ' What is x? ' flourishes ; and this question-form flourishes precisely because it is vague. It saves us the trouble of thinking out and saying exactly what we want to know about x . . . We can use this question-form to express a general desire to be given any useful information about x of any sort." The conclusion to be drawn, as Burt would no doubt agree, is that the editor of the 1921 Symposium asked an ambiguous question; and it is therefore scarcely surprising that he received a wide range of different answers. Burt's unexplained distinction, however, between "What do you conceive intelligence to be?" and "How is intelligence to be defined?" seems to me to confuse his main argument rather than add to its effectiveness.

---

* *Loc. cit.*, p. 190.

Burt himself claims to be asking two questions: " . . . (i) how precisely should the term (intelligence) be defined, and (ii) what evidence is there for believing that something really exists corresponding to the definition proposed?" In what sense of 'definition,' then, is the formula "innate general cognitive ability" intended to be a definition? How much can we infer from his general argument about what is required of a "definition in the strict logical sense"?

"How *the term* should be defined." This remark suggests that Burt is concerned on the face of it with words rather than things. This does not, of course, mean that he is concerned *merely* to say that in sentences where the word 'intelligence' occurs the words "innate general cognitive ability" can be substituted (or should be allowed to be substituted). He assumes that we are already familiar with the words 'innate,' 'general,' 'cognitive,' and 'ability,' and can apply them correctly. His point is that we *need a word* to refer to the ability in question.

"The definition proposed." This suggests a proposal, a stipulation. It is clear, however, that Burt is not offering simply a stipulative definition out of the blue. He backs his proposal by what may be called "the appeal to Cicero." The argument on *p.* 160 of his paper seems to be "Cicero used the Latin word 'intelligentia' as a technical term. My stipulation involves no radical departure from Cicero's usage. It is not, therefore, a misleading Humpty-Dumpytism, like using 'glory' to mean 'a nice knock-down argument';* but my definition is in part lexical as well as stipulative."

To assess the merits of a lexical definition we examine its historical accuracy. To assess the merits of a stipulative definition we ask, Is such a stipulation useful? In the case of Burt's definition we are therefore required to do both.

Although it is outside the scope of this paper to carry out such a programme in detail, one comment may perhaps be made on the lexical side. Burt says on p. 160 that 'intelligence' is not a "word of popular speech." But this surely does not follow, as Burt seems to suggest, from the fact that Cicero used the Latin word 'intelligentia' as a technical term. There can be no reasonable doubt that many Englishmen use the word 'intelligence' without having the least idea how Cicero used it; some, indeed, may not have heard of Cicero at all. If the argument is to do what Burt wants it to do, he should surely indicate that his definition conforms not with Cicero's usage, but with the usage of a twentieth century English-speaking person. And it is surely precisely because the word 'intelligence' *is* a word of popular speech (and, therefore, 'open,' imprecise, ambiguous) that an accurate account of its use in popular speech cannot be given. It cannot be given not because none of us is clever enough to find the right formula, but because there is no formula to be found.†

---

* See *Through the Looking Glass*, Chapter VI: "'When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean—neither more nor less'."

† A further criticism of the historical section of Burt's paper should be made in passing, namely his interpretation of the Greek words θυμός and θυμοειδές. Burt suggests that these words have something in common with the English word 'moral'; but the interpretation 'moral' seems to me, not a "rough but inexact equivalent," as Burt says, but a definite blunder.

In any case, as has been indicated already, the whole question of whether a definition of intelligence conforms to ordinary usage is unimportant. Ordinary usage may be vaguely suggestive of possible test-items, but there is nothing more to it than that. A person who is dissatisfied with a definition of intelligence on these grounds can always invent a new technical term in its place. Burt, of course, would agree with this. As he makes clear later in his paper, he would be quite prepared to allow that in place of " innate general cognitive ability " some other label could be substituted.

The crucial problem, then, has nothing to do with conformity to ordinary usage. It is rather whether the policy commended by Burt is a helpful one to follow for future research. In a sense, what Burt is offering is a stipulative definition—a proposal how the word ' intelligence ' should be used. But he is doing more than this. He is, I think, offering a definition of intelligence in the two other senses which we mentioned earlier. He is offering (*a*) description plus naming, and (*b*) a key for understanding a wide range of subject matter. As regards (*a*), Robinson puts the matter in general form by saying " The character yz occurs, and I call this character (or it is commonly called) by the name x." Applied to this case, the formula becomes " Innate general cognitive ability occurs, and I give it the name ' intelligence '." As for (*b*), the key, according to Burt, is that we should *look for* an innate general cognitive ability. Later in his paper he is even more specific ; he has in mind a characteristic transmitted by the genes. The psychologist who studies intelligence must, in other words, become a geneticist ; and, if we direct our research in this direction, then according to Burt, we are likely to produce worthwhile results. Such claims, it should be added, are seldom decisively confirmed or refuted. Either they turn out to be fruitful and win general acceptance, or they gradually fall into oblivion.

Thus, (i) it would be no argument against Burt's policy to accuse him of " measuring something called ' intelligence ' without knowing what that something is or how it is defined " (Richmond, quoted by Burt, *loc. cit.*, p. 158). Burt makes quite explicit both what his policy is and what he is trying to measure. (ii) It is no argument against Burt that a person who does not know what ' intelligence ' means would be even less likely to know the meaning of " innate general cognitive ability," for he is not attempting to explain the meaning of the word to someone who does not know it already. (iii) For the same reason, it serves little purpose to discuss whether or not his definition conforms to ordinary usage—whether, for instance, in ordinary usage the word ' intelligence ' is taken as referring to something innate.

On the other hand it *would* be an argument against Burt's definition if all discoverable tests consistently failed to produce appropriate distributions—i.e., distributions that would be expected on the assumption that what is being measured is something transmitted by the genes. Thus, if identical twins reared apart regularly produced widely different scores, if changes in environment were regularly followed by large-scale differences in score, the notion of an *innate* ability would become progressively more uncomfortable. Again, if reduction of emotional tension (e.g., by psycho-analysis or some other method) regularly resulted in improved performance, the whole distinction between the ' cognitive ' and the ' affective ' sides of human nature might seem to be an unhelpful one, and the idea of a specifically ' cognitive ' ability just no use.* Burt's definition would then be useless in much the same way as it would be useless having a word—say 'hexothippus'—meaning "a horse with six ears." It is not that 'hexothippus' does *not* mean "horse with six ears" if we choose to make it mean that; it is simply that we do not need the word. We do not need it for the obvious reason that horses do not have six ears. Thus it is crucial for Burt's argument, as he himself realises, to establish that "something really exists corresponding to the definition proposed" (p. 159).

---

* This statement requires qualification. Correlation coefficients can be worked out between as many different abilities (or ' substrates ') as we please ; but no figures, whether high or low, can lead us to abandon a substrate which we are sure is helpful or accept one which we are sure is unhelpful. To work out correlation coefficients is to *operate* our hypothesis, not to *test* that hypothesis. Whether we are ' getting anywhere ' with our hypothesis has to be settled rather by the general ' feel ' of the research undertaken.

This brings us to a problem that is central in any theoretical study of intelligence—the problem of how to establish assertions of the form "There really exists an x" when 'x' is not a straightforward concrete noun. Whatever the practical difficulties, we all know how to set about establishing the existence or non-existence of giraffes, dodos, fairies, etc.; but whether a particular sort of *ability* exists is clearly a very different question from whether a particular sort of giraffe exists; and it is this that constitutes the central problem. Part of the proof of the 'real existence' of intelligence in Burt's sense depends on factorial analysis; and the question which this technique forces upon us is, What conditions must be satisfied if factorial analysis is to justify statements of the form "There really exists an ability (or a factor) so and so"? This question can, I think, be satisfactorily answered as follows. It is not a sufficient condition for asserting the real existence of a factor that correlation coefficients between tests should form this or that pattern. A factor has a 'real existence' only when it becomes identified—that is, when the results which produced the particular pattern of correlation coefficients can be linked with independently discoverable events. Spearman (1927) has a chapter entitled "Proof that g and s exist"—his 'proof' being to show that, under suitable conditions, the surprising result of vanishing tetrads can be obtained. This is a misleading use of the word 'exist'. According to the usage which I would commend, "vanishing tetrads occur" is *not* a sufficient condition for asserting "g and s exist"; some independent attempt to identify g is also required. In this respect Spearman's suggested identification of g with 'mental energy' seems scientifically useless, since it suggests no independently discoverable events beyond those which in fact *did* produce the vanishing tetrads. In contrast, the factors of both Burt and Thomson (1950) have at least a chance of having a 'real existence,' for there is a suggested link between the test-behaviour which produced particular correlation coefficients and the independently discoverable behaviour of genes or neurones. There is a further possible justification for speaking of the 'reality' of factors. A factor is real, it might be said, if from behaviour at one test successful prediction is regularly made about behaviour at another test allegedly saturated with the same factor. This, however, opens the door to the postulation of a host of useless factors. We can be most sure of progress when links are found, not between one test and another, but between a person's test behaviour and the behaviour of genes and neurones. Genes and neurones, be it noted, are parts of the *body*; and, to put the matter epigrammatically, we could perhaps say that factors have a real existence, not when they are "factors of the mind" (whatever that means), but when they are *factors of the body*. Burt's definition involves the hope that the postulation of such genes will lead to a wide range of important findings. If this hope is fulfilled, his definition will have justified itself; otherwise not.

VI.—REFERENCES.

BRIDGMAN, P. W. (1927). *The Logic of Modern Physics.* Macmillan.
BURT, C. L. (1947). *Mental and Scholastic Tests.* Staples.

BURT, C. L. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.*, 25, 158-177.
HEIM, A. (1954). *The Appraisal of Intelligence.* Methuen.
PIAGET, J. (1950). *The Psychology of Intelligence.* Routledge and Kegan Paul.
ROBINSON, R. (1950). *Definition.* Oxford University Press.
RYLE, G. (1949). *The Concept of Mind.* Hutchinson.
SPEARMAN, C. (1927). *The Abilities of Man.* Macmillan.
THOMSON, G. H. (1950). *Factorial Analysis of Human Ability.* University of London Press.
WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins.

(*Manuscript received 2nd July,* 1956.)

II.—NATURE-NURTURE BALANCE SHEETS.

BY H. MADDOX
(*Education Department, University of Birmingham*).

I.—*The nature-nurture antithesis.* II.—*Attempts to partition test variance.* III.—*Twin resemblances.* IV.—*Interaction of nature and nurture.* V.—*Environmental influences.* VI.—*Summary and conclusions.* VII.—*References.*

SUMMARY.—Empirical studies have brought the nature-nurture controversy no nearer to a generally agreed solution. Attempts to derive a nature-nurture balance sheet from genetic theory fail to take into account the correlation between heredity and environment. Shuttleworth's widely quoted balance sheet rests on some doubtful assumptions. The only unequivocal evidence is derived from identical twins reared apart, but, in any case, any particular balance sheet can be neither final nor of general applicability since heredity and environment are in dynamic interaction. Test variance in groups whose education has been defective is likely to include an unusually large environmental contribution.

I.—THE NATURE-NURTURE ANTITHESIS.

THE British tradition in psychology has stressed the biological and hereditary determinants of behaviour to the relative neglect of social and cultural determinants. The Social Darwinism of Galton, Karl Pearson and McDougall, led to a line of statistical and ' factor ' psychologists who have tended to stress natural inequality. British psychologists have remained more attached to such concepts as ' innate ability ' than their counterparts in America where the *tabula rasa* doctrine—" the formation of all human character by circumstances, through the universal Principle of Association, and the consequent unlimited possibility of improving the moral and intellectual condition of mankind by education " (Mill, 1873)—was resurrected by the Behaviourists, substituting the principle of Conditioning for the principle of Association.

Sir Cyril Burt, for example, in his recent review (Burt, 1955) of the evidence for the concept of intelligence, reiterates his view that the word ' intelligence,' as a technical term in psychology, is best defined as " innate, general, cognitive ability." He is careful to distinguish between " intelligence as an abstract component of the individual's genetic constitution (γ) and intelligence as an observable and empirically measurable trait (g)." The " innate, general, cognitive ability " is the abstract ' γ ' not the observable 'g,' but he goes on to argue that " at least 75 per cent. of the measurable variance . . . is attributable to differences in genetic constitution, and less than 25 per cent. to environmental conditions." His review thus resuscitates the controversy concerning the relative contributions of nature and nurture to test variance. Burt himself states that the present ' excessive ' emphasis on environment is the result of the ' swinging pendulum of fashion ' rather than of the discovery of new facts.

While it must be agreed that little new evidence bearing on the traditional formulation of the nature-nurture problem has appeared in recent years, it seems that an environmental emphasis (which is more characteristic of America than of Britain), has come about, not as Burt suggests, because of changes in fashion, but because investigators have failed to produce convincing evidence about the genetic components of intelligence. Environmental factors are, moreover, capable of control and manipulation in a way that hereditary factors are not, and for this reason are of greater interest to educators.

Interest in the nature-nurture problem has declined because investigators have been unable to frame experiments to settle their disagreements: the problem has come to be regarded as 'metaphysical' since hypothetical concepts such as 'innate ability' cannot be operationally defined. Vernon (1937), for example, wrote that "arguments about nature and nurture in intelligence testing seem anachronistic; for surely most psychologists are by now agreed that no test can measure innate intelligence except in so far as it is expressed through the medium of acquired experiences." It was long believed that, although clearly any particular test score represents the combined and inextricable effects of 'heredity' and 'environment,' the crucial separation of these factors could be made by analysing the scores of identical twins reared apart (heredity constant, environment variable) and of foster children reared in the same home (heredity variable, environment constant). The last major American Nature-Nurture survey, however—the Thirty-Ninth Yearbook (1940) of the National Society for the Study of Education (which included some confessions of failure even to summarise the massive literature on the subject) succeeded only in revealing the apparently irreconcilable differences between traditionalists such as Terman and Goodenough and the environmentalists such as Stoddard and Wellman. Indeed, one contributor to the volume, H. E. Jones, seems to have felt that potential contributors should now be advised that nature-nurture had now been shown to be an unproductive field of research.

If the traditional formulation heredity *versus* environment is retained, the necessarily inconclusive evidence tends to be interpreted according to politico-philosophic bias, as Pastore (1949) has shown. Similarly Hebb and Thompson (1954) say: "one group of studies has seemed to show that environment can have little to do with intelligence; another that it must have a great deal. In either case the argument is somewhat indirect and not clearly decisive, and the literature itself has not led to a resolution of the apparent contradiction." In this situation it is hardly surprising that modern American texts by-pass the problem as unprofitable, if not meaningless. Tests are then regarded simply as work samples predictive of scholastic performance and the hypothetical 'intelligence' is discarded. Cronbach (1949), for example, writes: "The Stanford-Binet test measures present ability, not native capacity. While it seems obvious that no test can measure anything but here-and-now behaviour, there has been much confusion in the past thirty years because intelligence was thought of as hereditary." The crude dualism heredity *versus* environment has now given way to attempts to partition test variance into hereditary and environmental contributions. This paper will examine some of the attempts that have been made to arrive at a balance sheet in this fashion.

II.—ATTEMPTS TO PARTITION TEST VARIANCE.

Fisher (1918) attempted to use the technique of correlation to ascribe to heredity and environment fractions or percentages of the total variance which they together produce. Fisher justified his procedure, with respect to stature, by the fact that the variances are additive when the several causes of variability are independent. It at once appears, however, that, with respect to intelligence, the causes of variability are not independent. 'Heredity' and 'environment' are correlated, or, in other words, intelligent parents provide intellectually stimulating environments for their children, and *vice versa*. As Hogben (1939) points out, Fisher's partitioning procedure assumes that each genotype has an equal chance of experiencing each of the available kinds of environment, whereas "the viviparous habit of the human species, the existence of the family as a social institution and the stratification of society in widely different social levels all conspire to create a framework of environment which is intimately related to the distribution of genetic differences."

This objection is not decisive, however. It is possible to allow for the correlation between heredity and environment. Shuttleworth, for example (1935), computed that 16 per cent. of the I.Q. variance in Burks' study should be attributed to the correlation between heredity and inter-family environmental differences. He further estimated that where the environmental and hereditary contributions were of approximately equal magnitude, this 'interaction' term would account for approaching 50 per cent. of the variance: children would resemble their parents in intelligence, not primarily because superior parents pass on superior intellectual endowments, but because superior parents provide a higher order of intellectual stimulation and a healthier environment. Thus, a nature-nurture balance sheet which does not take account of this correlation will overstate the contribution either of heredity or of environment.

In his analysis of variability in stature, which Burt suggests is applicable to mental test data, Fisher concluded that "it is very unlikely that so much as 5 per cent. of the total variance is due to causes not heritable." But, as Hogben pointed out, his estimate was based on comparisons between observed correlations between relatives and correlations calculated from purely genetic considerations. In simple terms the genetic argument is that the correlation of ·5 with parental intelligence underestimates the influence of heredity, since only one-half of the chromosomes of each parent are present in each child. Thus, to the contributions of the two parents should be added a further contribution for atavism. As about 50 per cent. of the variance is still unaccounted for, however, it is necessary to invoke other factors such as segregation and assortative mating in order to avoid attributing the balance to environmental factors. As suggested above, however, the initial step in this argument is itself suspect, because of the correlation between heredity and environment. Indeed, family correlations in social attitude data have been found to resemble closely correlations in test data; but to account for family resemblances in social attitudes by heredity would clearly be absurd. It seems profitless, therefore, to pursue the genetic argument more fully.

Shuttleworth analysed Freeman, Holzinger and Newman's twin data and Burks' data concerning a group of foster children and a matched control group of children living with their true parents. He arrived at a nature-nurture balance sheet, which is quoted with approval by Vernon (1950) in his account of the effects of environment on test scores, as follows:

64 per cent. hereditary factors
16 per cent. environmental differences between families (inter-family differences)
3 per cent. differences in upbringing between children in the same family (intra-family differences)
17 per cent. joint heredity-environment factors, i.e., to the correlation between heredity and environment.

Shuttleworth's summary of his results is a little more complex, and applies to "California children of non-Hebrew, North European stock reared in the homes of their true parents . . ."

In order to obtain an estimate of intra-family variance the degree to which the resemblance in test scores of the identifical twins falls short of unity is used to estimate the effect of accidental and intra-family environmental influences. It has been demonstrated by Wilson (1934), however, that the environments of identical twins are more similar than the environments of fraternal twins, and on the face of it 3 per cent. of the variance for intra-family influences seems an underestimate. Indeed, Kelley (1930) arbitrarily decided that intra-family differences were as important as inter-family differences. It seems certain that for children of opposite sex and different birth order intra-family influences will be greater than for identical twins. Nevertheless, the variance of 3 per cent. is extrapolated and used in equations based on the foster children studies, and since in Shuttleworth's computation hereditary contributions are the residual percentage remaining after taking account of all other influences, his estimate of hereditary influences is exaggerated by the amount of the underestimate.

Shuttleworth's ingenious solution was made possible because Burks' control and foster groups were matched both for I.Q. variability and for environmental variability. He assumed that there was no selective placement of the foster children. The multiple correlation between I.Q. and a battery of environmental measures was ·42 for the foster children and ·61 for the control children. The square of the correlation for the foster children was used to obtain the percentage contribution of inter-family environmental influences (17·6 per cent.), and it was assumed that, since the groups were matched, the percentage contribution would hold good for the control of children also. It was then shown that for the control children 63 per cent. of the variance should be attributed to hereditary differences, whereas the remaining 16 per cent. was due to the correlation between heredity and environment. Thus, the final balance sheets for the two groups were (approximately) :

	Control Group	Foster Group
	%	%
Hereditary differences	63	79
Inter-family environmental differences	18	18
Intra-family environmental differences	3	3
Correlation between hereditary and inter-family differences	16	0

This at least confirms the point made by Haldane (1938) that there is no general solution to the nature-nurture question but only a series of particular solutions : the exceptional position of foster children results in abnormally large hereditary contributions to I.Q. variance.

III.—TWIN RESEMBLANCES.

In his article, Burt presents some new evidence on the resemblance of identical twins reared together and apart, non-identical twins reared together, foster siblings reared rogether and apart and children living with their parents. He emphases the high correlation in test scores between identical twins reared apart, ·843. This correlation, based on 21 pairs, is substantially higher than that reported by Newman, Freeman and Holzinger, ·670 for 19 pairs. In the American study the correlation for identical twins reared apart is very close to the correlation for non-identical twins reared together. In Burt's study, however, the resemblances between non-identical twins reared together are not much closer than those between ordinary brothers and sisters.

Although correlations based on such numbers are hardly conclusive, the closer resemblance of the identical twins reared apart would seem at first sight to justify Burt's emphasis on hereditary determination. In the American study, however, it is noteworthy that I.Q. differences between identifical twins reared

	BURT		NEWMAN, FREEMAN AND HOLZINGER	
	Identical twins reared apart	Non-identical twins reared together	Identical twins reared apart	Non-identical twins reared together
Individual test of Intelligence	n=21 ·843	n=172 ·526	n=19 ·670	n=52 ·640

apart depend partly on age at testing—the oldest and longest separated differing most—and partly on the degree of resemblance in the foster environments. Woodworth (1941) pointed out that the greatest pair differences occurred where there were marked differences in the environments of the twins, e.g., in rural—urban milieu and educational opportunity. Thus, relevant information which should accompany data on identical twins reared apart would include age at separation, age at testing, a description of the difference in environments and some estimate of the relative educational and social advantages of each member of the pair. In the absence of this information it might be assumed that Burt's pairs resemble each other more closely than the American pairs because pair differences in homes and schooling were less. In the American study the correlation of intra-pair differences in I.Q. with years of formal schooling was ·791. In the three cases in which there were the most marked differences in environment and schooling, the I.Qs. were 116-92, 106-89 and 96-77. Such findings illustrate that test performance is related to scholastic opportunity.

To summarise, in any attempt to distinguish between the contributions of heredity and environment to I.Q. variance, some account needs to be taken of the correlation between endowments and environments. The identical twins and foster children who are taken as the subjects for such study tend to be derived from a restricted range of environments. In any given society the range of environmental differences is likely to be more extensive than the range to be found in foster and adoptive homes. If, in the calculations, heredity is the residual category, it will tend to be inflated as a result. The balance sheet will tend to vary, depending on the amount of variation in both heredity and environment in the samples taken.

IV.—INTERACTION OF NATURE AND NURTURE.

Loevinger (1943) has pointed out that there is a good deal of evidence for the interaction of hereditary and environmental determiners. The term ' interaction ' is used in two different senses. When it is said that heredity and environment interact what is meant may be that there is a correlation between heredity and environment—that they are not distributed randomly with respect to each other. This is the sense in which Shuttleworth uses interaction when he refers to the correlation or ' joint effect ' of heredity and environment. In statistical discussions, however, interaction is used in a second sense, namely to describe what occurs when the effect of one factor is dependent upon the value of another factor. Brownlee (1947) insists that, in order to get a fair assessment of the effects of varying a particular variable, other related variables must be allowed to vary over their full range as well. Any linear assumption in nature-nurture analyses implies that a given grade of environment produces a given increment regardless of the grade of heredity raised in that environment.

Hogben gives a number of examples from experimental biology to show that different environments produce different effects on different stocks. Some breeds of chicken, for example, breed true for yellow shanks and some breed true for white shanks. The appearance of yellow shanks in the phenotype, however, is dependent on an adequate supply of xanthophyl in the diet. Thus, phenotypically, in environment A both stocks would be indistinguishable as regards this character whereas in environment B the yellow shank stock would clearly be different from the other. From such considerations Hogben concludes that where variations in environmental conditions produce differential effects on different stocks, the question of how much of a given effect is due to heredity and how much due to environment is devoid of meaning. Thus, it is safe to speak of genetic difference between two groups measured in one and the same environment or of a difference due to environment when identical stocks are measured under different conditions of development; but how much of a specific phenotype is due to heredity and how much to environment is a meaningless question since some alleles determine the same phenotype under all external conditions, while others may produce different phenotypes under different external conditions.

It is doubtful whether children can ever be said to have the 'same' psychological environment. Individuals in the same home, although experiencing more or less identical physical environments, have differing roles and social relations within the family. Thus, since the psychological environment cannot be held constant, studies of foster children cannot be said to hold 'environment' constant. The only unequivocal information is that derived from studies of identical twins reared apart in contrasting environments. Neither can studies of the degree of resemblance of identical twins as contrasted with the degree of resemblance of fraternal twins furnish evidence of 'hereditary determination.' The environments of identical twins are more similar than the environments of fraternal. Nevertheless, Blewett (1954), in a study of the resemblances of 12-15-year-old identical and fraternal twins in Thurstone's Primary Mental Abilities, concluded that 'Verbal Comprehension,' among other units of the test, was largely determined by heredity—against the great body of evidence that verbal tests at these ages are, of all tests, most influenced by environmental factors.

In summary, since mental measurement necessarily deals with the phenotypical expressions of intelligence in particular environments, and since individual performance is influenced by environment to an unknown degree, it is a meaningless question to ask how much of performance is due to heredity and how much to environment, except in the special case of identical twins reared apart.

V.—ENVIRONMENTAL INFLUENCES.

Having concluded that the nature-nurture question is virtually unanswerable, it may seem illogical to go on to illustrate situations in which the effect of environment is unusually great; but, since test score is still often equated with innate ability, it may be as well to summarise some studies of the influence of environment on (*a*) racial differences; (*b*) class differences; (*c*) underprivileged groups; and (*d*) post-school measurements.

In an earlier article, Burt (1943) states that, "from Binet onwards practically all the investigators who have attempted to construct 'intelligence tests' have been primarily searching for some measure of inborn capacity, as distinct from acquired knowledge and skill, and the I.Q. (corrected or uncorrected) should be used as the best available measure of the individual's inborn ability." While this is historically true and still seems to be current practice in educational psychology in this country, the history of testing in the U.S.A. has led to very different conclusions.

(*a*) *Racial differences.*—The early controversy over white-Negro intelligence led to some convincing demonstrations that the automatic equation of test score and inborn ability was unwarranted. This led to the further conclusion that since it was pointless to compare the test scores of Southern Negroes and Northern Whites, it might also be pointless to compare test scores of different socio-economic groups within the same society (Neff, 1938). In the group tests of World War I it was found that the average mental age of Negroes was 10·4 years, of whites 13·1 years, with an overlap betseen the two groups of only 12 per cent. For a decade these results were generally accepted as evidence of the intellectual inferiority of Negroes. (At this time likewise sociologists compiled much evidence on the pathological features of Negro life, which was interpreted as demonstrating the biological inferiority of Negroes.) In the thirties, however, this interpretation of the data was questioned. It was pointed out that some of the tests used were highly verbal, and that the vocabulary tests and wording of the instructions would be particularly difficult for Southern Negroes and for the foreign-born and their children. Such a language handicap was obviously not congenital and the tests could, therefore, not be considered measures of innate ability. Differences between white and coloured groups in motivation were also suspected.

When the scores of whites from the Southern 'Deep South' were tabulated against scores of the few Negroes from the industrial North, the Negroes were superior. The possibility that the Northern Negroes were unusually intelligent through selective immigration appeared unlikely when Klineberg (1935) demonstrated that the mean verbal I.Q. of Negro immigrant children in New York was increased by seven points by 3-4 years' residence in the city, and that newer arrivals from the South scored slightly higher than earlier immigrants. From such findings it was concluded that tested intelligence was at least partly a function of scholastic opportunity, and this conclusion was extended to explain not only racial differences but rural-ruban and social class differences in test performance.

(*b*) *Class differences.*—It has been repeatedly demonstrated that there is a correlation of about ·4 between I.Q. and the occupational class of a child's parents. A mean I.Q. difference of some 20 points has been found in many studies between the children of professional parents and the children of unskilled labourers. There is little agreement about the interpretation of these differences however. According to Eells (1951), factors which may contribute to 'status' differences in intelligence test performance include : (*a*) genetic ability ; (*b*) delevopment factors—such as degree of environmental stimulation ; (*c*) cultural bias in test items which are drawn from cultural materials unequally accessible to testees ; (*d*) test motivation ; and (*e*) test work habits or test skills. In an item analysis of 647 items in commonly used tests Eells showed that about half of the items in the tests for 9-10 year-old pupils and about 85 per cent. of the items for 13-14-year-old pupils showed significant differences between high and low status groups. Eells, admitting that his findings were open to other interpretations, concluded that " variations in opportunity for familiarity with specific cultural words, objects or processes required for answering the test items were the most adequate explanation of most of the findings." Perhaps the most common reply to accusations of class bias in test items has been the somewhat cynical statement that Western society is permeated with middle-class goals and values anyway, and that to take away the 'class' content of test items would rob them of much of their predictive value in a middle-class society. For purposes of straight scholastic prediction, the educational and verbal content of a test is no handicap. Verbal tests, however, correlate rather more highly with status than other tests, and hence so-called culture-free tests have their uses in investigations in which it is desired to minimise status influences on test score.

(*c*) *Underprivileged groups.*—That lack of normal educational opportunity depresses test scores, at least on such tests as the Binet, is indeed, obvious when we consider the verbal and scholastic content of many of the Binet items, and the fact that items become increasingly abstract and verbal at the higher ages. On the other hand, there is evidence that, given normal schooling, there is no cumulative age effect of social class on Binet scores: Terman and Merrill (1937) show that the differences in mean I.Q. between children from their highest and lowest occupational groups are almost as great at ages 2-5½ as at ages 15-18 years.*

All studies agree in finding an age decrement in I.Q. in children with restricted educational opportunity. Gordon (1924) noted that the mean I.Q. of canal-boat children on the 1916 Stanford-Binet was 90 at 5 years and 60 at 12 years with a negative correlation of − ·76 between age and I.Q. Among gypsy children he found a negative correlation of − ·45 between age and I.Q. Although Gordon's samples were small, similar results have been obtained by others. Asher (1935) reported the mean I.Q. of Kentucky mountain children as 84 at 7 years and 60 at 15 years. Edwards and Jones (1938) found the mean I.Q. of Georgia mountain children 108 at 7 years and 70 at 15 years. Similar age decrements have been reported for children educated in one-room rural schools (Anastasi and Foley, 1949). Many observers have noticed that children from underprivileged groups lose interest in school and sometimes become sullen and intractable when they become aware of the social consequences of minority group membership (Maddox and Morey, 1950).

Thus, whatever nature-nurture balance sheet might hold good for ordinary children in ordinary environments, it seems likely that for children from underprivileged groups, who figure disproportionately in the work of psychologists and social workers, the environmental effect on test scores will be substantial, particularly with older children.

(*d*) *Post-school measurement.*—The measurement of intelligence in post-school years shows even more convincingly that test scores cannot be used as measures of innate ability. A person who has spent a decade in a white collar occupation will clearly be at a great advantage in a paper and pencil test over one who has spent a decade in manual work. It may, of course, be objected that white-collar skills involve the exercise of intelligence, but as Wechsler (1944) observes, there are other kinds of intelligence besides the intellectual ability measured by paper and pencil tests, such as social and practical intelligence.

On leaving school the amount of subsequent schooling affects verbal I.Q. In a Swedish study by Husen, quoted by Anastasi (1955), for example, group intelligence tests were administered to 613 20-year-old recruits, all of whom had been tested at the age of 10 years. Amounts of schooling which varied widely in the sample, correlated ·61 with initial test score and ·80 with final rest score. Vernon and Parry (1949) reported that, while spatial and mechanical abilities appeared to rise between the ages of 14-18 years with or without schooling, verbal and educational abilities drop unless stimulated by secondary schooling, although '*g*' remained fairly constant.

---

* Their sample was, however, strongly biased in the direction of superiority of occupational status, only 6·6 per cent. falling into the lowest occupational class, as against an estimate of 10·5 per cent. of the total U.S. population in this class in the 1930 census. Thus, the underprivileged groups—the 'lower-lower' class—were inadequately represented in the sample.

A striking demonstration of the effect of environment on test scores was Tuddenham's (1948) cross-standardisation of the American Army Alpha and the AGCT. He found that the median soldier of World War II fell at the 83rd centile of the World War I population. This could only be explained by an increase in education and test skills in the population at large in the inter-war years.

In summary, the results of intelligence testing in a number of fields are inconsistent with the interpretation of test score as a measure of 'innate ability.' This is perhaps most clearly seen from the results of groups whose education has been defective.

VI.—SUMMARY AND CONCLUSIONS.

1.—The nature-nurture antithesis, for long a field of partisan controversy, has not been brought any nearer a generally agreed solution by empirical studies. Disagreements appear to be based chiefly on personal philosophies and unstated assumptions which control the interpretation of facts. For this reason many investigators appear to have decided that the traditional formulation of the nature-nurture problem is fruitless.

2.—Attempts to partition test variance into hereditary and environmental contributions need to take into account the correlation between heredity and environment. This would appear to rule out balance sheets based on deductions from genetic theory. Shuttleworth's widely quoted balance sheet has been shown to involve some doubtful assumptions.

3.—Studies of identical twins reared apart yield the only unequivocal evidence. It seems unlikely, however, that adequate controlled studies of this kind are feasible.

4.—In any case, the logical objection that heredity and environment are not disjunctive factors but in dynamic interaction, would rob any particular solution of its finality.

5.—A stock formula for the relative contributions of heredity and environment would be particularly misleading if applied to underprivileged groups. The cumulative effect of environment in such groups invalidates the interpretation of tests scores in terms of innate ability.

6.—The practical consequence of taking test scores as the best measure of innate ability is that those who are not sophisticated enough to maintain the distinction between 'intelligence' and 'tested intelligence' may use 'innate ability' as a first and sufficient explanation of poor performance ; whereas poor performance, particularly that of older underprivileged children in verbal tests, for example, may result equally from any of the five factors listed by Eells. Since genetic ability is the only one of the five factors which cannot be changed, inherent brightness or dullness should only be used as a last resort when other possible explanations have been ruled out.

VII.—REFERENCES.

ANASTASI, A. (1955). Individual differences. *Ann. Rev. Psychol.*

ANASTASI, A., and FOLEY, J. P. (1949). *Differential Psychology.* Macmillan.

ASHER, E. J. (1935). The inadequacy of current intelligence tests for testing Kentucky mountain children. *J. Genet. Psych.*, 46, 480-486.

BLEWETT, D. B. (1954). An experimental study of the inheritance of intelligence. *J. Ment, Sci.*, 100, 922-933.

BROWNLEE, K. A. (1947). *Industrial Experimentation.* London : H.M.S.O.

BURT, C. (1943). Ability and income. *Brit. J. Educ. Psuchol.*, 13, 83-98.

BURT, C. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.*, 25, 158-177.

CRONBACH, L. J. (1949). *Essentials of Psychological Testing.* Harper.

EDWARDS, A. S. and JONES, L. (1938). An experimental and field study of North Georgia mountaineers. *J. Soc. Psychol.*, 9, 317-333.

EELLS, K., *et al.* (1951). *Intelligence and Cultural Differences.* Univ. of Chicago Press.

FISHER, R. A. (1918). The correlation between relations on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.*, 52, 399, 433.

GORDON, H. (1924). Mental and scholastic tests among retarded children. *Bd. Educ. Pamphlet No.* 44. London: H.M.S.O.

HALDANE, J. B. S. (1938). *Heredity and Politics.* (Allen and Unwin).

HEBB, D. O., and THOMPSON, W. R. (1954). The social significance of animal studies. In Lindzey, G. (Ed.) *Handbook of Social Psychology.* New York: Addison-Wesley.

HOGBEN, L. (1939). *Nature and Nurture.* Allen and Unwin.

KELLEY, T. L. (1930). The inheritance of mental traits. In Murchison, C., *Psychologies of* 1930. Worcester: Clark Univ. Press, 423-443.

KLINEBURG, O. (1935). *Negro Intelligence and Selective Migration.* New York: Columbia Univ. Press.

LOEVINGER, J. (1943). On the proportional contributions of differences in nature and nurture to differences in intelligence. *Psychol. Bull.*, 40, 725-756.

MADDOX, H. and MOREY, D. (1950). *Dark People in Melbourne—A Study of Aboriginal Half-Castes.* Melbourne: Victorian Council of Social Service.

MILL, J. S. (1873). *Autobiography.* Longmans, Green.

NEFF, W. S. (1938). Socio-economic status and intelligence: a critical survey. *Psychol. Bull.*, 35, 727-756.

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION (1940). *Intelligence: Its Nature and Nurture, Part I.* Bloomington, Ill.; Public School Publishing Co.

PASTORE, N. (1949). *The Nature-Nurture Controversy.* New York: Columbia Univ. Press.

SHUTTLEWORTH, F. K. (1935). The nature versus nurture problem. *J. Educ. Psychol.*, 26, 561-578 and 655-681.

TERMAN, L. M. and MERRILL, M. A. (1937). *Measuring Intelligence.* Harrap.

TUDDENHAM, R. D. (1948). Soldier intelligence in World Wars I and II. *Amer. Psychologist*, 3, 54-6.

VERNON, P. E. (1937). The Stanford-Binet test as a psychometric method. *Char. and Pers.*, 6, 99-113.

VERNON, P. E. (1950). Psychological studies of the mental quality of the population. *Brit. J. Educ. Psychol.*, 20, 35-42.

VERNON, P. E. and PARRY, J. B. (1949). *Personnel Selection in the British Forces.* Univ. London Press.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins.

WILSON, P. T. (1934). A study of twins with special reference to heredity as a factor in determining differences in environment. *Human Biology*, 6, 324-354.

WOODWORTH, R. S. (1941). Heredity and environment. *Social Science Research Council Bulletin*, 47.

(*Manuscript received* 29*th June*, 1956.)

III.—A STUDY OF COGNITIVE ERROR THROUGH AN ANALYSIS OF INTELLIGENCE TEST ERRORS.

BY P. J. LAWRENCE

(*Lecturer in Education, Canterbury University College, Christchurch, New Zealand*).

I.—*Introduction.* II.—(*A*) *Systematic theories of cognitive error.* II.—(*B*) *Types of error noted in experiments on reasoning and problem solving.* III.—*Nature of the investigation.* IV.—*Results.* V.—*Discussion.* VI.—*References.*

SUMMARY. 1.—Descriptions of the processes underlying intelligence have moved away from formal statements such as Spearman's principles of noegenesis towards looser statements centred on reasoning or problem solving skills.

2.—An analysis of intelligence test errors, in order to throw some light on the operation of these skills, should, therefore, be carried out within the context of the more general study of cognitive error.

3.—Spearman and Selz have developed systematic theories of cognitive error. Both direct experimental evidence, and indirect evidence arising from the types of error reported in the literature on problem solving and reasoning, favour Selz's ' task displacement ' rather than Spearman's ' reproduction.'

4.—From a group of 274 children (median age 11 years 8 months, mean I.Q. 104), whose errors on the Otis group verbal test had been analysed quantitatively, a sample of 84 was chosen for more intensive qualitative analysis. This analysis was based on the re-working of all items which were originally wrongly answered.

5.—Conditions producing error were classified as : inadequate formulation of the task ; prepotency of particular details or characteristics ; mis-reading, or failure to read all the choices offered ; correct working but incorrect expression of the answer ; difficulty in ' recentring ' ; lack of confidence ; correct answer but for a wrong reason and rejection of the correct answer on reasonable grounds ; lack of specific information ; complete failure to formulate the task.

6.—These results are interpreted as supporting the type of explanation offered by Selz and paralleling the conditions noted in studies of problem solving and reasoning. The main implication suggested is that teachers should have a greater concern for the development and exercise of the problem solving skills of children as an essential element in the development of intelligence.

I.—INTRODUCTION.

ERRORS may be studied from at least two viewpoints—as providing information which can be used directly to improve performance through remedial teaching or coaching, or as leading, indirectly, to a clearer understanding of underlying processes. This paper is concerned with the latter viewpoint, in particular the thesis that a study of the errors made in an intelligence test should throw some light on the nature of the processes underlying intelligence.

These processes have rarely been defined, psychologically, as carefully as the various ' factors ' have been defined statistically ; thus many descriptions of the nature of intelligence have been marked by looseness and generality. Spearman's principles of noegenesis (particularly the eduction of relations and correlates) have often been accepted as representing a systematic description of the qualitative aspect of intelligence.*

---

* Writers from quite different standpoints, e.g., Knight (1948) and Rimoldi (1951), have drawn heavily upon these principles in their interpretations of intelligence.

In his last major paper, Spearman added ' abstraction ' to noegenesis and claimed that these together describe the psychological nature of ' g ' (Spearman, 1946).

On the whole, however, there has been a reluctance to formulate precise definitions of intelligence, and a tendency to identify it with traditional but loosely defined cognitive processes. This can be illustrated by the tentative definition given by Hearnshaw (1951, p. 315) : " . . . a cluster of high grade skills concerned with problem solving," and by Vernon's statement (1950, p. 54) which sums up a widely accepted viewpoint, i.e., " . . . reasoning ability is one of the commoner definitions of intelligence." Burt (1955), in discussing briefly the cognitive nature of the general factor suggested that its characteristic functions may be represented by such concepts as integration and direction. Among those who prefer a multi-factor interpretation, there is the same emphasis on the central role of reasoning, e.g., Thurstone (1941) found that the inductive factor had the highest correlation with the second order general factor, and in a later survey (1946) suggested that the inductive factor along with two closure factors (strength and flexibility of closure—described in Thurstone, 1944) might represent the central intellective factor. Guilford has also emphasised reasoning, e.g., " . . . the two aptitude scores representing verbal comprehension and general reasoning cover fairly well the variance in traditional intelligence tests " (Guilford and Lacey, 1947, p. 55).

In order to clarify the nature of the processes underlying intelligence, it is apparently necessary to analyse more closely the fundamental operations involved in reasoning or problem solving skills. It is not the primary purpose of this paper to make such an analysis.* But, in order to interpret intelligence test errors adequately, it is necessary to select one special area of investigation, i.e., cognitive error. This falls into two parts (A) an evaluation of systematic theories of cognitive error, and (B) a survey and classification of types of error reported in analyses of problem solving and reasoning.

II.—(A) SYSTEMATIC THEORIES OF COGNITIVE ERROR.

Spearman has been specifically mentioned above, because he has contributed one of the few systematic accounts of the nature of error. A contrasting analysis is that of Selz, these two accounts both springing from general cognitive theories.

(i) *Spearman.*—In an article on the origin of error, Spearman (1928) claimed that the only sound theory from which error can be predicted is his theory of noegenesis, in particular as it is affected by one of the minor quantitative laws, i.e., retentivity. He summed up his position thus : " . . . the liability to error must needs ensue whenever an item arises in any context by virture of reproduction, but is then taken as being there by virtue of intuitive evidence such as invests the three principles of noegenesis . . . " (p. 34). He did not concede the possibility of incorrect eduction being a cause of error ; in fact although he pointed out that the ascribing of error to creative imagination would be a formidable objection, he quickly dispensed with the objection by stating that it is the process of mistaking the mental presentation for real existence which is the error, not the actual educing of the correlate.

A further example sums up this point of view and illustrates its weakness : " . . . any item or relation whatever may fit their original context perfectly well and yet, on being retained when this context changes may result in presenting any degree of misfit " (p. 33). He emphasised the condition of ' being retained,' yet it would seem that the error arises from failure to understand the implications of the change of context, while reproduction of the item or relation arises *as a result of* this faulty approach to the problem, i.e., the crucial factor is relevance, not retentivity.

---

* An attempt to do this through integrating findings from a number of studies in related cognitive fields, has been made by Lawrence (1955, pp. 7-61).

(ii) *Selz.*—A critical account of Selz's theory of thinking has been given by Humphrey (1951), who while dubious of some of the major aspects of the theory, yet stressed the originality and importance of the emphasis on ' total integration ' in thinking. This process of integration is illustrated by the *Aufgabe* or task complex, which initiates and directs the finding of the means to the solution. Selz's specific theory of error, derived from this emphasis on the task complex, has been described in detail by Wilcocks (1928).

The theory is based on the assumption that errors arise, not because of strong reproductive tendencies, but as a result of the particular task-directed intellectual operations which serve the solution. " The most common ground for these errors was, in his experiments, that a related problem served instead of the one actually set " (Wilcocks, p. 27)—this may be called ' task displacement.' Conditions which may produce such displacement appear to be dependent upon a law of part efficiency, i.e., not all parts of the task actually set are sufficiently active. This law operates through the dropping of some characteristics of the task, the prepotency of particular elements, the low dispositional effects of some elements and readiness to attend to the more abstract or general characteristics of the task.

(iii) *Relevant experiments.*—There have been few experiments bearing directly on these two theories. Wilcocks (1928) set out to examine a number of introspective protocols of simple reaction tasks, in order to produce evidence which would show that Selz's analysis does not exclude the process of ' substitution '—a process essentially similar to Spearman's reproduction. The evidence suggested that ' substitution ' and Selz's ' part efficiency ' are not mutually exclusive processes, but may both operate to cause error.

One of the great limitations of Selz's work was that he based his experiments on a very narrow type of situation, e.g., such problems as : " Give the co-ordinate of parson." In view of this limitation, an experiment by Fortes (1932) in which the errors on perceptual intelligence tests were analysed in the light of Selz's theory, is significant. His subjects were children of an average age of 13 years, and his analysis was based on objective records checked by questioning " a large number of children," although he did not give any further details about this questioning. In Fortes' classification of the errors found, reproduction or perseveration accounted for very few, while the main emphasis was on ' displacement,' i.e., displacement of cognitive energy from a total situation to a part only. He suggested that the antecedent condition for the occurrence of such errors was the indeterminate cognition of the total perceptual situation.

Brigham (1932) made an extensive study of items used in College Scholastic Aptitude tests, one of his major findings being that errors are " characteristic solutions of the problems set." Spearman's theory was rejected as merely a labelling of the one process noegenesis if the solution happens to be right, and reproduction if it happens to be wrong. He pointed out that, although the problems may be presented and the answers recorded simply, yet the psychological structure of the situations may be exceedingly complex. The displacement of a problem from a narrow to a much broader (and mainly irrelevant) context is illustrated in his observation that the group refused to accept the printed premises as true, but judged them " in the wider universe of experience."

On the whole his attitude is summed up in the statement " . . . our studies have indicated error to be not capricious, but orderly " (p. 195).

From these few studies bearing directly on the two major theories, it is apparent that Spearman's theory of reproduction is inadequate, while concepts based upon the subject's perception of the ' task complex ' (Selz) have proved more fruitful in explaining the origin of error.

II.—(B) TYPES OF ERROR NOTED IN EXPERIMENTS ON REASONING AND PROBLEM SOLVING.

In many experiments on reasoning and problem solving, specific types of error have been noted. Although these come from studies with different purposes, subjects and materials, there are enough regularities to suggest classification in a few broad categories.

(i) *Inadequate formulation of the problem.*—Piaget (1924) pointed out that a mistake may be the result of " a faulty point of view, due to the fact that he has not yet asked himself the question as we ask it " (p. 90), i.e., there is some formulation of the problem, but not an adequate formulation. Other reports of this type of error are : Chrisof (1939), failure to formulate the problem, or an " inability to penetrate the occasion " ; Eagleson (1940), insufficient knowledge of directions ; Smith and Tyler (1942), failure to study the description of the problem fully enough to appreciate all the limiting factors ; Bloom and Broder (1950), confusion about the requirements of the problem ; Burak (1950), incomplete formulation of the task ; and Jones (1953), failure to see the problem clearly or to consider all its aspects.

This type of error exemplifies what Selz emphasised, i.e., the response is incorrect because the task is not clearly structured. It also supports the distinction between eduction of relations (or correlates) in general, and *relevant* eduction in a particular context.

(ii) *Distortion or partial operation of the task.*—This is really a special case of the first category, and again Piaget (1926) has illustrated its essential characteristic. " Syncretistic understanding consists precisely in this, that the whole is understood before the parts are analysed, and that understanding of the details takes place rightly or wrongly, only as a function of the general schema " (p. 151). One of the most penetrating accounts of the way in which such a schema operates has been given by Bartlett (1932) in his studies of perceiving and remembering, and his observations on levelling, sharpening, conventionalizing, and emphasising have been confirmed in studies of problem solving.

Maslow (1948) has noted the re-shaping of a problem as the result of assumptions about what is involved, while Hildreth (1941) has described " the difficulty reduction tendency "—i.e., the subject *remakes* a problem in an effort to make it meaningful.

(iii) *Dislocation of the elements in a problem.*—Thorndike (1917) used the terms " dislocation or disrelation of elements "—a failure due to an inadequate analysis of the elements in a problem. This has been noted specifically in problem solving situations, e.g., by Chant (1953), where the ' analytic ' approach was replaced by an ' intuitional ' approach, and Burak (1950), who found that failure to analyse out major variables was one of the important criteria of unsuccessful solution. Speed has been listed by Eagleson (1940) as a condition leading to failure in analysis, while Maslow (1948) has pointed to the potency of the unusual and unfamiliar in distorting analysis.

The prepotency of particular elements as a disturbing factor has been demonstrated by Woodworth and Sells—Woodworth (1938), Hildreth (1942), Morgan (1944) and Simmel (1953)—studies which illustrate not only the attraction of prepotent elements, but the effects of overlearning of certain ways of perceiving data.

(iv) *Influence of ' set.'*—Overlearning may produce a set, and apart from the major studies of Maier and Luchins (see below), set or stereotyped response has been reported as a cause of error in a number of studies. For example, Chrisof (1939) reported inhibition of the clear meaning of the task because of a particular set ; Bloom and Broder (1950) noted failures due to an inability to change a relationship once it had been established ; and Fattu, *et al.* (1954) found that the best criterion for distinguishing the poor from the good problem solver was ' stereotypy of behaviour.'

Another condition which is related to this general type of error is the weakening or loss of the ' task set,' i.e., losing track of the problem after having started out in the right direction. Bloom and Broder (1950) described this as losing sight of the plan of the problem.

(v) *Influence of attitude, emotional involvement and personal preference.*—Such conditions as maladjustment (Staudt, 1949), emotional involvement (Marks, 1950, and Marks and Ramond, 1951), low level of aspiration and threat (Combs, 1952) have been reported as leading to a greater proneness to error. Bloom and Broder (1950) found that the particular format of a problem was sufficient to discourage students from attempting any attack, while some adopted an all-or-none attitude, i.e., " either I can do this type of problem or I cannot do it." Eagleson (1940) reported a situation which is commonly recognised in school work, but which is perhaps too often unrecognised in testing, i.e., lack of interest. The influence of strong personal attitudes or prejudices in causing error in syllogistic type problems has been noted by several investigators, e.g., Brigham (1932) gives many examples.

(vi) *Gestalt account of error.*—Although the assumption is made throughout Gestalt writings, that failure to structure or re-organize may lead to error, yet there are a few specific references to error worth noting. Duncker (1945) mentioned error due to attempts to escape from the original setting of the problem, ' single track ' proposals which leave out of consideration all other factors, and confusion of the original problem with another for which the correct solution is known. Similar sources of error were noted by Wertheimer (1945), i.e., inadequate view of the whole situation, lack of breadth of view—" Even when he has it at the beginning he may lose it in the process because he is busy with details," over-extended view, premature or short cut closure—" Often the impatient desire to find the solution focuses the eye too exclusively, too strongly " (p. 195).

Errors arising from the ' blind ' transfer of habits have been described in detail by Maier (1933) and Luchins (1942), as ' habitual direction ' and ' mechanisation,' the one resulting from the typical experience of the individual with the materials of the problem, the other being created by the particular mode of arrangement or presentation of the materials.

While it is not possible to fit all these types of error into one consistent theory, it is apparent that many of the conditions described indicate a failure to appreciate the total structure of the task—a characteristic more in keeping with the theory of Selz than with that of Spearman.

III.—NATURE OF THE INVESTIGATION.

(i) *Purpose.*—The general purpose was to examine the errors made in an intelligence test in order to throw more light on the characteristic processes involved in answering test items. Types of error observed could then be compared with those noted above, drawn from studies of problem solving and reasoning, and with the contrasting theories of Spearman and Selz.

(ii) *Test and subjects.*—The test chosen for analysis was the Otis Intermediate Form B group verbal test—a test which is typical of the many verbal tests which have been used for both educational and experimental purposes.* Furthermore, the Otis is made up of a number of different types of items, thus providing scope for the interplay of many facets of problem solving and reasoning processes.

Subjects were drawn from all the Form I classes of an intermediate school, N=274, median age 11 years 8 months at time of entry. The mean I.Q. was 104·3, S.D. 12·6, just prior to entry and 107·7, S.D. 13·4 on a second testing seven months later. As far as could be judged, the sample was representative of urban Form I children, and provided a relatively homogeneous age group (interquartile range of 12 months), with a wide range of intelligence. This particular age group was chosen in order to meet two criteria: (*a*) the children should be old enough to cover most of the items, thus giving scope for the production of as many errors as possible; (*b*) on the other hand, they should not be at the top of the age and class range for the test, as this would increase the likelihood of the ' ceiling ' effect being too widespread.

(iii) *Method of analysis.*—It has been customary to analyse results from a large number of subjects working on many items, by calculating for each item the frequency with which particular wrong answers are given, a method used by Hamid (1925), Brigham (1932), Fortes (1932) and Eells, *et al* (1951). This method was adopted for preliminary analysis in the present study, the results confirming the general finding that errors are very unevenly distributed in both multiple choice and open-answer type questions.† Thus, many of the items had 50 per cent. or more of the errors concentrated on one particular distractor or wrong response.

It is legitimate to use such a method for selecting items for further study, in the light of some specific criterion, e.g., socio-economic status differences in error responses, as in the Eell's study. It is highly questionable, however, whether it is legitimate to *infer* from the frequency of a particular error that a certain process has operated to produce this response. The most direct method of investigation is the actual re-working, by the child, of the incorrect items, and this method was chosen as being most likely to give the fullest information about the conditions leading to the production of error.

In preliminary trials it was found that merely asking the child to work the question aloud was quite inadequate, and that it was necessary for the experimenter to ask supplementary questions, e.g., " why did you decide to do that . . . or choose that ? " ; " what was puzzling you about this question . . . where does it tell you what to do . . . which is the important part of the question . . . what is the meaning of this word ? " These questions were also asked occasionally on items where there were no errors, so that the children would not associate them with failure. There were considerable individual differences, at all I.Q. levels, in the facility with which the children could explain how they arrived at particular answers.

---

* This test has been standardised for N.Z. children and has been widely used in N.Z. schools, particularly as one of the criteria for the placing of children in ability and attainment streams in Intermediate schools, i.e., the last two years before secondary schooling.

† This analysis, along with other exploratory analyses of test errors, is described more fully in Lawrence, P. J. (1957) : Some characteristics of incorrect responses to intelligence test items, *Aust. J. Psychol.* (in press).

From the total of 274, a sample of 84 children was used for questioning, 6 boys and 6 girls from each of 7 I.Q. groups (123+ ; 117-122 ; 111-116 ; 105-110 ; 99-104 ; 93-98 ; 92 or less). Apart from these limitations, the children were chosen at random. All children had taken the Otis test twice—the November before entering intermediate school (schools begin in February) and early in the following July, while the questioning was carried out in November and December following the second test. Originally, it was intended to limit the questioning to only those items which were wrong on both tests. However, in the preliminary trials it was found that this resulted in the loss of much valuable information, so that finally all the errors on both tests were re-worked. The gap between the second test and the questioning was intentional; for if the same error was still made on re-working five months later, then it suggested some stability or consistency in the underlying condition responsible for the error.

Interviews took, on the average, 1¼ hours, with a maximum of 1½ hours and a minimum of 25 minutes for some of the brightest children. As the children were familiar with the experimenter, from his having been in the school for some time giving a series of tests, satisfactory rapport was easily established, and most children were very keen to co-operate.

IV.—Results.

Verbatim records of the children's explanations and answers to the experimenter's questions, for each error made on both tests, were grouped into eleven categories according to the type of item involved.* This made it possible to compare the classifications of the condition giving rise to error, in groups of items of different content and structure. Limitation of space makes it impossible to reproduce in detail the classification of these conditions for each category; but this is not a serious curtailment, for the major conditions noted ran through almost all the categories, despite the very different content and structure of the items.

Because of the combination into broad groups of similar conditions found within different categories, and also because of the frequent overlap of two or more conditions in the one response, estimates of the percentage of the total errors falling into each group are necessarily of limited value. An overall figure has, however, been given for each major group of conditions as summarized below.

(i) *Inadequate formulation of the task* (31 per cent.).—This was the most frequent cause of error, over all I.Q. groups, and pointed clearly to a tendency for children to under-estimate the necessity for being clear as to exactly what is given, and exactly what is required. Sometimes it took the form of the expansion of a specific requirement into a more general requirement, e.g., in Analogy items a general relation " this *goes with* that " was used, rather than a specific relation, " this is a *particular function* of that." From the explanations and questioning it was clear that many children were quite capable of educing the correct relation and then finding the correlate, but they did not stop to consider the implications of the exact form of the question. A typical comment was: " I just wasn't bothering much about the first part of the sentence."

---

* The seventy-five items in the test were classified into these eleven categories—word meaning, opposites, word or sentence attribute or definition, forming a concept from given examples, mixed sentences, classifying and arranging elements of a concept in a given order, analogies, verbal reasoning, proverbs, number series, arithmetical reasoning.

In other cases, unfamiliar or complicating aspects of the problem were ignored, from one word up to half of the item—a type of ' difficulty reduction.' Examples of this are found in the responses to the item: " Write the letter that follows the letter that comes next after K in the alphabet." Children from all I.Q. groups gave ' L,' having ' reduced ' the item to " write . . . letter . . . follows . . . K." The record of one child (girl I.Q. 116), illustrates the stubbornness with which some clung to their original interpretation. When asked to read aloud, she still gave ' L ' ; again, after reading slowly, she gave ' L,' and could not see anything wrong with her answer until the item was read for her twice, with emphasis.

Another illustration of inadequate formulation is the reformulation or reconstruction of the item to fit in with an assumption about what is required. Typical cases are shown in the interpretation of the item: " If a man has walked east from his home 7 miles and then walked west 4 miles, how many miles is he from home? " In some cases the words " from his home " were added after " west," in others the last part of the item was changed to " how many miles has he walked? " Once such assumptions had been made, even re-reading slowly aloud often failed to shift them. The substitution of a completely different task, due to perseveration of methods used in earlier items, was sometimes responsible for inadequate formulation, e.g., transferring the set for finding opposites to items which called for the application of quite different relationships.

(ii) *Prepotency of particular details or characteristics of the task* (8 per cent.). Although this condition is really a variant of (i) above, it had specific characteristics which warrant separate illustration. One such characteristic was the losing of ' direction,' or the establishing of a false ' direction,' while concentrating on particular details of the data. A typical example was the response of one boy (I.Q. 121) to the item: " Which one of the following words comes last in the dictionary?—heart, judge, grass, nerve, horse, north, labour." He picked out the two competing words, i.e., nerve and north, compared the second letters and said: " e and o to choose between, well e comes before o so the answer must be nerve." He had established a directional set " which comes last? " but as a result of concentration on details, this set is weakened and the habitual alphabetical direction of " e before o " displaced it.

Many errors of this general type could have been avoided if a check had been made before writing the answer, in order to make sure that the broad requirements of the task had been met. It was obvious, however, that in most cases the production of at least some sort of answer satisfied the child, and verification or checking was rarely attempted.

Narrowing or distorting the task by concentrating on particular details, or a particular characteristic, is a related condition. One item read: " Which one of the five things below is most like these three: cannon ball, wire, penny?—pound note, bone, string, pencil, key." Narrowing the field of search, which prevents the acknowledgement of the correct response, is illustrated by the record of one girl (I.Q. 128) who gave ' string ' as her answer—" because a cannon ball is round, a penny is round, and wire looking at it from the end is round." When asked if there was anything else similar about them she said: " . . . they are all made of copper . . . so could be ' key,' but *there are all shapes of keys*." Although she was close to grasping the essential characteristic, having decided on ' roundness ' she could not break way from her concept of ' shape.'

In some cases the correct answer was passed over, even though fully understood, in favour of a distractor with particularly prepotent qualities or associations. The most extreme instances of this condition were those in which a particular distractor was chosen *because of* its unfamiliarity, even though the relevance of the correct answer was understood. One boy (I.Q. 101) when explaining his choice, said that he thought that the answer was likely to be one of the unfamiliar words which he did not know, so he chose the most unfamiliar, despite a good understanding of the correct answer. A typical illustration comes from a response to the item : " Which word means the opposite of succeed ?—win, decline, fail, accede, try." A girl (I.Q. 110) explained that she saw ' accede,' did not know it, so thought it must be the right one—although she understood the relevance of the correct answer. A related attitude, namely the belief that there must be a ' catch ' somewhere—that the item is really meant to be more complicated than it looks was illustrated by the boy (I.Q. 101) who explained that he was not taking account of the first part of the question, " . . . because the first part might be just to fool you."

(iii) *Failure to read or to consider all the choices offered* ; *mis-reading, and inability to read a word understood aurally* (14 per cent.).—This condition again illustrates a feature of the test attitude of many children, namely a failure to distinguish between a relation of general relevance and the specific relation appropriate to the item. When re-reading and re-working items, children often gave evidence that, having chosen an answer of general relevance, they did not even look at the other choices to see whether there was anything of more specific relevance offering. Many spontaneous changes to the right answer occurred when, for the first time, all the available choices were considered.

This condition is illustrated in responses to the following item : " An event which is sure to happen is said to be (?) probable, certain, doubtful, possible, delayed." A boy (I.Q. 121) who corrected his answer on re-reading, explained his original choice of ' probable ' by saying " I took the first one that looked something like it." A girl (I.Q. 103) who did the same thing, explained : " I saw ' probable ' so did not look over the others." On a different type of item, another boy (I.Q. 94) changed to the correct answer spontaneously and remarked that he decided to look over the words again, as he had not really looked at them carefully before.

As well as the ignoring of part of an item or the misinterpreting of a crucial phrase in an item, there was also simple misreading. Examples were—adding words to the original statement, which complicated or simplified the task ; reading ' brag ' as ' brat ' or ' bag ' ' mayor ' as ' major,' ' instruction ' for ' institution,' ' succeed,' as ' suicide ' or ' ceased ' etc.

With some of the duller subjects quite familiar words were mispronounced and therefore meaningless, although when these were pronounced for them, the item caused no great difficulty.

(iv) *Correct working of the problem, but failure to express the answer in the specific terms required by the instructions* (13 per cent.).—This condition, equally prevalent at all I.Q. levels, was not merely due to a hasty reading of the items. Even in the reasonably leisurely conditions of the interview, children persisted in their faulty formulations of otherwise correct answers. Instructions within mixed sentence items such as " with what letter would the last word of the sentence begin ? " . . . or in number series items such as " what should that (i.e., the correct) number be ? " were persistently ignored, distorted, or forgotten. In some cases re-reading aloud was ineffective, and it was necessary to ask specifically about what was actually required before the child appreciated his faulty expression of the correct answer.

(v) *Difficulty in ' recentring,' i.e., a lack of flexibility in shifting a word from one content to another, or in breaking up a word combination with a strong ' gestalt '* (7 per cent.).—An example of the former was the inability to interpret the word ' sum ' as the total of an addition, once it had been thought of in its general connotation of " a problem in arithmetic." Similarly ' letter ' as a written communication often inhibited the alternative interpretation of a unit in the alphabet. A strong combination which caused great difficulty was ' honey bees '—many children, on the particular mixed sentence item in which this occurred, attempted to rearrange the other words in ingenious ways to fit in with this pair. Once having paired them they could not make the necessary separation for correct rearrangement. Not only did these difficulties set the children off in the wrong direction in attempting to rearrange the words, but they resulted in sets which were strongly resistant to any restructuring.

(vi) *Lack of confidence leading to avoidance of the item, or over confidence leading to satisfaction with an incomplete or poor solution.*—No percentage has been given for this group because, in the re-working, avoidance was ruled out as the child was expected to attempt each item ; thus any subsequent error was classified under the other groups. An interesting example of avoidance is illustrated by the fact that of all the omissions made by the total group from which this sample of 84 children was drawn, 42 per cent. were on one only of the eleven categories, i.e., mixed sentences, and 14 per cent. on a similar type of item, i.e., classifying and arranging elements. It would appear from these percentages that, because these items were cast in an unusual, and at first sight confusing, form, children just omitted them. This was confirmed by the manner in which the individual children reacted to these items, expressing dislike and a lack of confidence in their ability to solve them. When urged to try, many children found that they could tackle the items and rearrange the words in the correct order. Similar expressions of lack of confidence were made on items in categories such as verbal reasoning, number series and arithmetical reasoning, by particular children who showed, however, that they were, in fact, quite capable of working them correctly.

In some cases a wrong answer was given with an over-confidence which inhibited any attempt to check over the item, and despite the fact that the child knew it to be a poor or inadequate solution. There were many contrasts in the children's approaches to the items—from the very cautious response to the rather carefree gamble.

(vii) *Correct answer, but with wrong or inadequate reason ; the reading of a wrong but not completely excluded interpretation into the data ; rejection of the correct answer on reasonable grounds* (15 per cent.).—At all I.Q. levels there were many correct answers given with wrong or inadequate reasons—in some cases due to a superficial or devious chain of reasoning, in others due to the particular structure of the item. The second and third conditions listed above are of more significance, however, because they indicate how difficult it is to present the child with a clear-cut problem, that is, clear-cut from the child's point of view. This was particularly evident in items of the type " Which tells best just what a lamb is ? " Children were often confused on such items because they could not decide from whose point of view ' tells best ' should be interpreted. The correct answer ' a young sheep ' was rejected for various reasons, e.g., one girl (I.Q. 119) argued that people might not know what a sheep is but there is more likelihood that they would understand the description ' an animal with wool ' ; a boy (I.Q. 125) qualified his correct choice after much hesitation—" . . . but only if you know already what a sheep is " ; another boy (I.Q. 102) commented " . . . if you didn't know what a sheep was you wouldn't know anything " (i.e., about a lamb).

These are typical of many genuine hesitations due to some confusion as to what the item really means. This weighing up of the meaning and intention of the item was summed up in the comment of one boy who acknowledged, somewhat reluctantly, the reasonableness of the correct answer : " . . . if you look at it that way " !

(viii) *Lack of specific information required for the problem* (8 per cent.).—This condition was most evident in the verbal categories, e.g., word meaning and, in particular, proverbs. For example, in items beginning " There is a saying . . . " success seemed to depend to a considerable extent on previous experience with ' sayings.' Children at all I.Q. levels had confused ideas as to what a ' saying ' was although, when it was explained to them, many were able to tackle the items sensibly.

As would be expected, vocabulary deficiencies were greatest in the lower I.Q. groups, although not confined to them. It was noticed that where a little known word was introduced into an item (e.g., malicious, institution) the effects were unpredictable. Some children ignored the word completely and yet chose the correct answer *in spite of it*, others while being attracted to the correct answer, rejected it *because of* the unknown word.

(ix) *Complete failure to formulate the task* (4 per cent.).—This condition did not appear as often as it would have if all the children in the lower I.Q. groups had completed the more difficult items at the end of the test. On the whole, however, there was a tendency for the tasks to be at least partially formulated, as the children attempted to extract some meaning from the items. The distinction was not a clear cut one between understanding and not understanding the nature of the task ; rather there were degrees of understanding, dependent upon the degree of adequate formulation of the task.

V.—DISCUSSION.

(i) *Comparison with cognitive errors in general.*—These results provide little evidence to support Spearman's theory of error ; in fact, the items most likely to show the operation of reproduction indicate that failure is due to the educing of an inadequate relation, rather than to a failure to educe at all. An example, which on superficial inspection would seem to support Spearman, was found in responses to the item—" Grass is to cattle as bread is to what— ? butter, flour, milk, man, horses." In the total Form I group 101 errors were made on this item, 53 per cent. on ' butter ' and 41 per cent. on ' flour.' This could be interpreted as a clear indication of reproduction due to a strong association displacing eduction ; but, in the individual analyses the explanations, pointed to an interpretation based upon the child's inadequate formulation of the task, i.e., of what was really required. The explanation of one girl (I.Q. 96) summed up a typical attitude—she chose ' butter ' because " . . . it is the nearest to bread—you have to find *what goes with it.*"

On the whole, Selz's concepts of task displacement and part efficiency prove to be of considerable value in interpreting the origin of error, although this does not entirely rule out Spearman's analysis. Such conditions as the carrying over of a distorting set, or the transfer of inappropriate methods or assumptions, clearly involve reproduction. What the evidence does suggest is that reproduction may operate as one particular process resulting from the more basic condition of inadequate formulation of the task—thus Selz's interpretation contains, yet goes beyond that of Spearman.

Fortes's analysis of errors can be matched quite closely with the analysis presented above, and this is all the more significant as his study was not based on verbal, but on 'perceptual,' intelligence tests. One general comment of Brigham's was amply illustrated in this analysis, namely that problems were displaced " from a narrow to a much broader (and mainly irrelevant) context."

Of more significance than the relevance of Selz's theory to the detailed conditions described is the close similarity of these results to the conditions of error noted in diverse experiments on reasoning and problem solving. This similarity of process confirms the value of regarding intelligence as a composite of specific cognitive skills.* These skills appear to depend not only upon innate conditions which set a limit to the complexity of what may be perceived and related, but also upon the cumulative effect of the numerous and varied learning experiences of the child. It is through these experiences that skills are developed and made readily available.†

As pointed out in the first paragraph of the Introduction, the classifications of error presented could be used as a basis for coaching, in order to obtain at least short-term increases in I.Q. A more fundamental implication is that teachers should not adopt a fatalistic attitude towards intelligence, but should appreciate that, from a long-term point of view, the educational experiences of the child have a significant bearing on the formation and use of the specific cognitive skills underlying the manifestation of intelligence.

Test constructors attempt to eliminate some sources of error by such improvements as better setting out of items, less ambiguous wording and more practice examples.‡ It is apparent, however, from the results of this study, and the experience of reworking items with children, that there is a more significant source of error—namely the firmly established work habits, attitudes or assumptions which children bring to the test situation. These are likely to have been overlearned, having been reinforced in many varied situations during the school years, and thus to exercise a constant and important influence on the approach to any cognitive task.

Inadequate formulation of the task, to take an example, is not just the result of undue haste or a lapse in attention; it is also a failure in a crucial cognitive skill—clarifying the requirements of a given problem.§ Such skill is developed, with more or less efficiency and conscious application, from the earliest years through a wide range of media. Attention to detail, verification or checking, flexibility in approach, consciousness of degrees of relevance rather than merely right *versus* wrong—these cannot be left to develop adequately without deliberate guidance and exercise. While teachers may normally accept this fact in relation to the particular subjects which they teach, it is suggested as a result of both the review of cognitive error and the specific investigation included in this paper, that they should be alert to the influence of such guidance on the develpoment of intelligence.

---

* In a recent article Guilford (1956, p. 170) writes: " We have come to the view that a common unique, *psychological* core for all problem solving does not exist. Problems are simply too varied, and each type seems to call upon its own pattern of abilities . . . "

† Both Hebb (1949) and Piaget (1950) have formulated theories in which an attempt is made to show how experience interacts with innately determined processes, in the development of intelligence. Hebb's " intelligence B," the average level of performance or comprehension, would represent the functioning of the cognitive skills referred to above—Cf. Burt's (1955) ' g,' i.e., intelligence as an observable and empirically measurable trait.

‡ The Otis test used in this study has since been revised by the A.C.E.R., with clearer setting out and the elimination of some unsatisfactory items.

§ Cf. Duncker (1945), explication of the data and analysis of the goal, or Wertheimer (1945), being sensible to the structural properties of the problem. Guilford (1956, p. 171) confirms the importance of this skill in his statement: " . . . *general reasoning* has something to do with comprehending or structuring problems of certain kinds in preparation for solving them."

VI.—REFERENCES.

BARTLETT, F. C. (1932). *Remembering.* Cambridge University Press.

BLOOM, B. S. and BRODER, L. J. (1950). Problem solving processes of college students. *Suppl. Educ. Mon.* University of Chicago Press.

BRIGHAM, C. (1932). *A Study of Error.* N.Y. College Entrance Examination Board.

BURAK, B. (1950). The nature and efficiency of methods of attack on reasoning problems. *Psychological Monograph No.* 313.

BURT, C. L. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.,* 25, 158-177.

CHANT, S. N. F. (1933). An objective experiment in reasoning. *Amer. J. Psychol.,* 45, 282-291.

CHRISOF, C. (1939). The formulation and elaboration of thought problems. *Amer. J. Psychol.,* 52, 161-185.

COMBS, A. W. (1952). Intelligence from a perceptual point of view. *J. Abnorm. Soc. Psychol.,* 47, 662-673.

DUNCKER, K. (1945). On problem solving. *Psychological Monograph No.* 270.

EAGLESON, O. W. (1940). A study of puzzle solving. *J. Psychol.,* 9, 259-268.

EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., and TYLER, R. (1951). *Intelligence and Cultural Differences.* University of Chicago Press.

FATTU, N. A., MECH, E., and KAPOS, E. (1954). Some statistical relationships between selected response dimensions and problem solving efficiency. *Psychological Monograph No.* 377.

FORTES, M. (1932). A study of cognitive error. *Brit. J. Educ. Psychol.,* 2, 297-318.

GUILFORD, J. P. and LACEY, J. I., eds. (1947). *Printed Classification Tests.* U.S. Govt. Printing Office.

GUILFORD, J. P., KETTNER, N. W., and CHRISTENSEN, P. R. (1956). The nature of the general reasoning factor. *Psychol. Rev.,* 63, 169-172.

HAMID, S. A. (1925). Some factors of effectiveness in mental (intelligence) tests. *Brit. J. Psychol.,* 16, 100-115.

HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.,* 42, 315-321.

HEBB, D. O. (1949). *The Organization of Behaviour.* New York : Wiley.

HILDRETH, G. (1941). The difficulty reduction tendency in perception and problem solving. *J. Educ. Psychol.,* 32, 305-313.

—— (1942). Puzzle solving with and without understanding. *J. Educ. Psychol.,* 33, 595-604.

HUMPHREY, G. (1951). *Thinking.* London : Methuen.

JONES, S. (1953). Process testing—an attempt to analyse reasons for students responses to test questions. *J. Educ. Res.,* 46, 525-534.

KNIGHT, R. (1948) 4th ed. *Intelligence and Intelligence Tests.* London : Methuen.

LAWRENCE, P. J. (1955). *The Significance of Method in Intellectual Tasks.* Unpublished Ph.D. Thesis, Canterbury University College Library, New Zealand.

LUCHINS, A. S. (1942). Mechanization in problem solving : the effect of Einstellung. *Psychological Monograph No.* 248.

MAIER, N. R. F. (1933). An aspect of human reasoning. *Brit. J. Psychol.,* 24, 144-155.

MARKS, M. (1951). Problem solving as a function of the situation. *J. Exp. Psychol.,* 41, 74-80.

MARKS, M. and RAMOND, C. K. (1951). A new technique for observing concept evocation. *J. Exp. Psychol.,* 42, 424-429.

MASLOW, A. H. (1948). Cognition of the particular and of the generic. *Psychol. Rev.,* 55, 22-40.

MORGAN, J. J. B. (1944). Effect of non-rational factors on inductive reasoning. *J. Exp. Psychol.,* 34, 159-168.

PIAGET, J. (1924). *Judgment and Reasoning in the Child.* London : Kegan Paul.

—— (1926). *The Language and Thought of the Child.* London : Kegan Paul.

—— (1950). *The Psychology of Intelligence.* London : Routledge and Kegan Paul.

RIMOLDI, H. J. A. (1951). The central intellective factor. *Psychometrika,* 13, 27-46.

SIMMEL, L. (1955). The coin problem : a study in thinking. *Amer. J. Psychol.,* 66, 229-241.

SMITH, E. R. and TYLER, R. (1942). *Appraising and Recording Student Progress.* New York: Harper.

SPEARMAN, C. (1928). The origin of error. *J. Gen. Psychol.*, 1, 29-53.

—— (1946). The theory of general factor. *Brit. J. Psychol.*, 36, 117-131.

STAUDT, V. M. (1949). The relationship of certain personality traits to errors and the correct responses in several types of tasks among college women under varying test conditions. *J. Psychol.*, 27, 465-478.

THORNDIKE, E. L. (1917). The psychology of thinking in the case of reading. *Psychol. Rev.*, 24, 220-234.

THURSTONE, L. L. (1944). A factorial study of perception. *Psychometric Monograph No.* 4.

—— (1946). Theories of intelligence. *Scientific Monthly*, 62, 101-112.

—— and THURSTONE, G. (1941). Factorial studies of intelligence. *Psychometric Monograph* No. 2.

VERNON, P. E. (1950). *The Structure of Human Abilities.* London: Methuen.

WERTHEIMER, M. (1945). *Productive Thinking.* New York: Harper.

WILCOCKS, R. W. (1928). On substitution as a cause of errors in thinking. *Amer. J. Psychol.*, 40, 26-50.

WOODWORTH, R. S. (1938). *Experimental Psychology.* New York: Holt.

Manuscript received, 27th August, 1956.)

IV.—INTELLIGENCE AND AGE : RETESTING WITH TWENTY-FOUR YEARS' INTERVAL.

BY JOHN D. NISBET
(*Lecturer in Education, University of Aberdeen*).

I.—*Introduction and Previous work.* II.—*Present study.* III.—*Results.* IV.—*Discussion.* V.—*References.*

SUMMARY. 1.—141 graduates (72 men, 69 women) tested on a shortened version of Simplex Group Test in 1930-34 at average age 22½ years, repeated the test in 1955 at average age 47, and scored 11.9 points of score higher on average on the later occasion.

2.—The improvement in test performance in the 24-year interval is estimated to amount to about 0·7 or 0·8σ in terms of the standard deviation in the whole adult population.

3.—The improvement in performance was significant in all but one of fourteen sub-tests, including both verbal and non-verbal material.

4.—The correlation between scores on the two occasions of testing was +0·48. When corrected for homogeneity, this represents a test-retest correlation of about +0·7.

5.—Part of the observed increase in score may have been due to (*a*) the difference in conditions of doing the two tests, and (*b*) general test sophistication ; but the whole of it could not be thus explained away.

6.—In view of similar findings from five parallel American investigations, also with subjects of advanced education and superior ability, it is concluded that such persons do not give evidence of decline in test performance at least up to the later forties.

I.—INTRODUCTION AND PREVIOUS WORK.

IN a previous paper (Nisbet, 1954), an account was given of a follow-up which tested the validity of intelligence test scores in predicting, among a group of graduates during their training as teachers, the level of responsibility of the teaching post held some twenty years later. The goodwill and co-operation shown in that investigation and the success in tracing 80 per cent. of the graduates suggested an extension of the inquiry to assess also the reliability of the test by means of a re-testing of the same group on the same test after this long interval. The results of this, while corroborating earlier results on test reliability from re-testing over shorter intervals, produced evidence on the relation between intelligence and age which ran counter to the general trend of results in that field.

It is not considered necessary to review here the substantial body of evidence which has shown a decline of intelligence with age once the years of childhood and adolescent growth are past. These results have been summarised by Wechsler (1944) :

> " Every human capacity after attaining a maximum begins an immediate decline . . . The age at which this maximum is attained varies from ability to ability but seldom occurs beyond 30 and in most cases somewhere in the early 20's . . . (p. 55).
>
> " Certain abilities decline more slowly with age than others. Thus, the abilities called for by the General Information and General Comprehension Tests hold up much better than do the abilities called for by the Substitution and the Memory Span for Digits Tests " (p. 63).

The evidence has been derived largely from standardization procedures for adult tests and from personnel selection both in industry and in the Forces. However, the method in these large-scale inquiries has been 'cross-sectional,' that is, involving the testing of different samples at different ages with careful checks to ensure that the various samples were closely equivalent. Since the exact equivalence of samples, especially at later ages, is difficult to guarantee, it is clearly important to check these results by a longitudinal study, testing a group during their alleged peak of ability and following up over a sufficiently long period to obtain evidence from this same group of the alleged decline in their ability with age. Such a method is difficult and time-consuming, but five investigations of this type have recently been completed: all of these show not a decline but an increase in average test score with age.

In 1952, Swanson reported the repetition of the Minnesota College Aptitude Test with a group of 55 men who had taken it as a college entrance test in 1926-28. Retesting after twenty years showed a substantial average gain, amounting to 35 points among the 22 who had successfully graduated from college, 9 points among the 13 who had attended college but not graduated and 7 points among the 20 who had not attended college. However, as the test is largely a measure of vocabulary, involving completions, synonyms and antonyms, perhaps this result is not surprising, as vocabulary is known to hold up well in maturity.

The monograph of Owens (1953) marks a substantial achievement in follow-up, for he retested 127 men on the Army Alpha Test thirty years after they had taken it as part of the entrance examination of Iowa State College in 1919. Some of the second testing was done at a reunion banquet and the rest by individual visits, where the subject received a five-dollar honorarium and the tester a 'comparable fee'; in this way 63 per cent. of the potential subjects were tested. The results showed a significant increase in mean score, the later mean score being about one-half of a standard deviation above the original mean. Significant increases occurred in the sub-tests, Practical Judgment, Synonym-Antonym, Disarranged Sentences, Information, and Analogies; increases were not significant in Following Directions and Number Series; and there was a decrease, not significant, in Arithmetical Problems. However, since the initial test in this case was done before entrance to college, Owens pointed out that the result does not rule out the possibility of decline with age: a substantial increase in ability as a result of college education might be cancelling the effect of a gradual decline subsequently.

In 1955, three further investigations involving retesting after long intervals in adulthood were reported. Bentz (1955) applied to 209 male executives of the Sears Company Chicago, the A.C.E.P. Examination which they had done about eight years previously. The published abstract of this paper presents a complicated result since the subjects were classified as high, average and low scorers on the basis of the first test score, and no allowance was made for the normal regression effect. Removal of this spurious influence shows (by calculation from data in the original paper) a small decline among those over age 40 in total score and also in both the verbal and problem-solving sections of the test, and a considerable increase among those under 34 in total score and verbal score, with a slight decrease (near zero) in the problem-section. His conclusion is that "the older group drops in ability . . . younger men consistently increase," their increase being "of such consistency and magnitude that one may view this age period" (20-34) "as one of intellectual growth."

Bayley and Oden (1955), working with the Stanford Study of the Gifted, report the testing of 768 'gifted' adults and 335 of their spouses with two forms of the Concept Mastery Test, which involves synonyms, antonyms and analogies, without time limit. The first test was at average age 29 and the second at average age 41. Equivalent scores for the two forms were established, and showed that " scores at the later testing . . . are consistently higher than for the first testing, the mean increase amounting to about half a standard deviation. The increases occurred in all occupational and educational levels represented, at all levels of ability tested, except where the test ceiling prevented, and at all ages from 20 to 50 years. The gains were found in both the synonym-antonym and the analogies halves of the test, but were somewhat larger in the former."

Finally Jones (1955) is reported to have compared performance on the Terman Group Test at ages 17 and 33, with 83 cases, and found an average gain of seven-tenths of a standard deviation. Such gain may of course have occurred wholly in the years immediately after 17, and re-test at 33 is rather early to be relevant to a discussion of decline with age.

These results suggest certain limitations on the general conclusions about the decline of ability with age. It is already accepted that verbal abilities hold up well with age, but the results show not merely the absence of decline of verbal abilities, but a marked increase, during the years from 20 to 40. Results on non-verbal abilities are not so consistent, and there is some doubt about the influence of the actual years of college education. It is to be noted also that the investigations are limited for the most part to adults of high ability and college education.

II.—Present study.

Group testing of post-graduate students training as teachers in Aberdeen was begun in 1930 by Professor Rex Knight.* The test booklets of the period 1930-33 fortunately escaped war-time salvage drives, and university records provided sufficient further information to allow a considerable proportion of these graduates to be traced in 1952. In many cases the original home address provided a means of contact twenty years later, and others were traced by devious means, particularly through the helpful co-operation of education authorities on whose staff they were or had been serving as teachers. After arduous inquiry, 266 (or 80 per cent.) of the 335 graduates for whom test records survived, returned forms giving their address and appointment and items of personal history. The responsibility of the teaching post held was assessed on a five-point scale ; and with this criterion of advancement in the profession, the intelligence test score showed a correlation of +0·26, whereas type of university degree (also on a five-point scale) gave a correlation of +0·36, and estimates of teaching ability during training gave zero correlations. These results were complicated by various factors which were discussed in a previous paper (Nisbet, *op. cit.*).

The results of this first stage of the inquiry were reported to the group individually. In 1955, the group was approached a second time with a request to repeat, in their own homes under strict but self-imposed conditions, the identical test which they had taken some twenty years previously. Of the 266, a considerable number had changed their addresses in the interval, letters being returned undelivered ; several had died ; and those resident abroad were excluded. 232 graduates were, in fact, approached with the request, and of these no fewer than 141 (or 61 per cent.) volunteered and carried out the test. These were subsequently notified of their own scores and also of the general results of the inquiry. The whole process required over 1,800 communications but produced in the end two sets of 141 test booklets, one set done in 1930-33 and the other done 22-25 years later by the same persons.

---

* The writer wishes to acknowledge the assistance of Professor Rex Knight and Dr. N. T. Walker, of the University of Aberdeen, who were responsible for the initial testing of the subjects and who generously gave permission for the use of the test results in this investigation.

The 141 volunteers (72 men, 69 women) were slightly but significantly superior in mean score on the first test to the total group of 232 approached. The mean score of the 141 was 68·0 (σ10·6); and of the 232, 64·9 (σ10·4). The proportion of men volunteering (72 per cent.) was higher than the proportion of women (53 per cent).

The mean ages on the two occasions of testing are given in Table 1. The range of ages was from 20 : 0 to 28 : 3 on the first test, with more than 90 per cent. below age 25 : 0. The range on the second test was from 43 : 0 to 53 : 4, with more than 90 per cent. over age 45. In no case was the interval between tests less than 21 years, and the maximum interval was 26 years, the mean interval being 24 years 6 months. In Table 1, the mean ages are presented separately for men and women and for ordinary and honours graduates, since the men tended to be older than the women and the honours course was one year longer than the ordinary course. For this reason, no attempt was made to relate age to score within the range of the separate testings, since any trend would be obscured not only by the differences between the lengths of courses but also by the fact that the more able students completed their courses at younger ages than the less gifted.

TABLE 1

MEAN AGES ON FIRST AND SECOND TESTING.

	Men			Women			Total Group
	Hons.	Ord.	All	Hons.	Ord.	All	
Mean Age on : 1st test	23 : 4	22 : 7	22 : 11	22 : 7	22 : 0	22 : 2	22 : 6
2nd test	47 : 11	46 : 11	47 : 4	46 : 9	46 : 7	46 : 8	47 : 0
No. of cases	33	39	72	14	55	69	141

The test used was a shortened version of the Simplex Group Test. The last fourteen sub-tests (Test 3B to the end) were set, with a time limit of 45 minutes without separate timing for sub-tests. These sub-tests are : Relations (verbal) ; Analogies (verbal) ; Instructions (verbal) ; Re-arrangement (verbal) ; Number Series ; Vocabulary ; Completion (verbal) ; Mirror Images ; Digit Memory ; Absurdities (verbal) ; Relations (verbal) ; Relations (number) ; Vocabulary ; and Substitution. On the first occasion the test had been done as part of the class work in Psychology. In 1955, when the subjects did the test in their own homes, a new front page was pasted on the test booklets, designed to ensure strict adherence to instructions and including a small flap, as in a theatre programme, which was gummed over to seal the booklet and discourage premature glancing through the test. Subjects were also asked to put the test into the envelope supplied and seal it immediately at the end of the time-limit.

III.—RESULTS.

Table 2 gives the mean scores in the two applications of the test.

In the group as a whole, the mean score in the second application of the test, at average age 47, was some twelve points higher than the mean score on the first testing at average age 22 : 6. The improvement with age is more than one-

TABLE 2

MEAN SCORES IN FIRST AND SECOND APPLICATIONS OF THE TEST.

	Men Hons.	Ord.	All	Women Hons.	Ord.	All	Total Group
Mean (1st test) ..	69·27	66·87	67·97	72·72	62·82	64·83	66·43
σ (1st test)	11·39	11·15	11·24	11·07	8·15	9·60	10·56
Mean (2nd test) ..	83·36	79·95	81·51	78·79	74·00	74·97	78·31
σ (2nd test)	10·10	10·56	10·42	9·53	9·11	9·32	10·40
Improvement	14·09	13·08	13·54	6·07	11·18	10·14	11·88
No. of cases	33	39	72	14	55	69	141

sigma in terms of the standard deviation of scores in the first testing. To express the amount of increase in this way is to over-estimate, since the graduates give a standard deviation smaller than that of a representative sample of the population. As the test used was a shortened version, no population norms for it were available. However, Scottish graduates are known on Moray House Adult Test I, a rather similar verbal test, to have a standard deviation of scores of approximately two-thirds that of a representative sample of the whole population ; and if this is assumed to apply also to the shortened Simplex used here, the amount of improvement is equivalent to about 0·7 or 0·8σ. This figure is close to those reported by Owens, Bayley and Oden, and Jones (*op. cit.*). In Owens' data, it was doubted whether a substantial part of the observed increase might not have occurred during college attendance at the very start of his thirty years' interval. The present results agree with those of Bayley and Oden in showing that even when the first test occurs after the years of college or university attendance, an increase in score is still found.

The improvement is significant at the ·01 level for both the men and women separately. There was a slightly greater improvement among the men than among the women ; but the difference between the increases of the men and of the women just fails to reach significance at the ·05 level (t=1·91).

Table 3 presents scores in the various sub-tests separately. Since the sub-tests vary in contribution to the total, some scoring six points and others twelve, some with wide scatter and others with little, the amounts of improvement in Table 3 have been made roughly comparable with each other by expressing the mean improvement in each case in terms of the standard deviation of scores in that sub-test on the first occasion of testing.

In every one of the fourteen sub-tests the score on the second occasion was higher than the score on the first. In every case except one, this improvement in score was significant. The one exception is explained by the fact that the average score in that sub-test on the first occasion was 5·1 out of a maximum of 6, and though the second testing showed a small improvement, the ' ceiling ' effect prevented it from reaching significance level.

It should be remembered that these sub-tests are not separately timed. This reduces their independence of one another in that a person who is able to complete rapidly, say, the verbal sub-tests thereby has more time to attack the non-verbal ones. There is, however, some justification for isolating the sub-test scores. Almost every person tackled the first eleven of these fourteen sub-tests

TABLE 3

MEAN IMPROVEMENT IN SUB-TESTS SEPARATELY.

Sub-test	Mean Improvement (in terms of σ)	Significance At ·05 x At ·01 xx	t-value
1. Relations (verbal)	0·39	xx	4·4
2. Analogies (verbal)	0·26	x	2·4
3. Instructions (verbal) ..	0·51	xx	6·0
4. Re-arrangement (verbal)	0·21	x	2·2
5. Number series	0·65	xx	7·1
6. Vocabulary	0·37	xx	4·4
7. Completion (verbal) ..	0·33	xx	3·3
8. Mirror images	0·40	xx	4·4
9. Digit memory..........	0·16	—	1·5
10. Absurdities (verbal) ..	0·32	xx	3·1
11. Relations (verbal)	0·33	xx	3·6
12. Relations (number)	0·44	xx	4·2
13. Vocabulary	0·67	xx	8·1
14. Substitution	0·71	xx	6·2

on both occasions and yet less than half had any time to spare at the end. The final sub-test was a time-absorber, a long code substitution test, slow and tedious, gaining relatively few points of score, and only 35 of the 141 subjects completed it on the first occasion and only 59 on the second. The effect of a single time-limit covering all fourteen sub-tests was thus probably not as great as it might be in other differently arranged tests.

Two further points remain to be mentioned in this comparison of performance on the two tests. Previous work has suggested that the effect of age on test performance is more marked on non-verbal and number items than on verbal ones ; and also that age particularly affects speed of performance, so that a strict time-limit unduly penalizes the older subject.

Table IV presents mean scores in certain sections of the test which were abstracted. Such abstraction of items from a test with an over-all time limit is of questionable validity, as mentioned above. The sub-tests included in the table are referred to by number, in accordance with the numbering in Table 3.

TABLE 4

MEAN SCORES IN VERBAL AND NUMBER ITEMS.

Type of item	Sub-tests included	Mean Improvement (in terms of σ)	Significance At ·01 xx	t-value
Verbal	1, 2, 3, 4, 6, 7, 10, 11, 13..............	0·86	xx	10·29
Vocabulary..	6, 13	0·70	xx	9·12
Number ..	5, 9, 12	0·66	xx	7·33

Both verbal and number items show significant increases in score. The amount of increase in a group of items, even when expressed in terms of the standard deviation of scores on that group in the first test, is to some extent influenced by the number of items in the group ; but allowing for this and for the influence of the over-all time limit, it would appear that the improvement is less marked in the number items than in the verbal items. There are too many assumptions involved to justify the application of precise assessment of the significance of the difference.

The relation of age to speed in test performance can be gauged from the proportions of subjects completing the test on the two occasions. As was mentioned earlier, 35 of the 141 subjects completed the whole test on the first occasion and 59 on the second. A similar result is obtained if the lengthy final sub-test is discounted : 43 of the 141 failed to start this final section in the first testing, and only 11 failed to start it in the second testing. The difference is highly significant ($X^2=23{\cdot}45$). Excluding this final test and the complicated number relations problem near the end, the average number of individual items omitted was 5·4 on the first and 2·6 on the second, again a significant difference ($t>4$).

The correlation between scores on the two applications of the test was +0·48, as shown in Table 5.

TABLE 5

TEST RE-TEST CORRELATIONS.

	r test-retest
Complete test, total group	+0·48
Ditto, corrected for homogeneity	+0·7
Complete test, men only	+0·52
Complete test, women only	+0·38
Vocabulary items (sub-tests 6, 13)	+0·48
Verbal items (sub-tests 1, 2, 3, 4, 6, 7, 10, 11, 13)	+0·44
Number items (sub-tests 5, 9, 12)	+0·39

Since the group is highly selected for intelligence, this correlation coefficient needs to be corrected for homogeneity. An estimate of the degree of selection in these scores on the shortened version of the Simplex Test was made, as described above, from approximately comparable data from Moray House Adult Test I. On this basis, the observed test-retest correlation of +0·48 would when corrected amount to about +0·7. Such a result is consistent with the findings of previous work on re-testing over shorter intervals, and represents a fair degree of agreement, with many small changes and a few large differences in score. The large differences were all improvements in the second test, five persons increasing their score by more than thirty points of score (32, 33, 34, 46, and 47 points). Only two persons decreased in score by more than ten points (12 and 14 points).

The test-retest correlation among the 72 men (+0·52) was higher than that among the 69 women (+0·38) ; but the difference between these coefficients was not significant ($P={\cdot}31$).

IV.—DISCUSSION.

The results thus agree with those of the five other previous studies in which re-testing was carried out well into adulthood, in showing an improvement in test performance. This finding from six longitudinal studies is at variance with well-established results from cross-sectional investigations on the relation of age and intelligence. However, it remains to be established in the present data that the improvement with age in test performance among the graduates is a genuine improvement in ability and not spurious.

There are at least two alternative possible sources of the observed increase in score, first, the difference in the conditions of doing the test on the two occasions, and second, general test sophistication. There is evidence to suggest that each of these may have contributed to the improvement in performance.

Firstly, there is the fact that the first test was done in a classroom under supervision as part of a course of professional training, and the second was done at home under self-imposed test conditions. The steps taken to assist strict observance of rules in the second test have been described above ; but there is always the possibility (if the subjects will excuse the remark) that some cheating, deliberate or accidental, may have occurred. However, cheating would have to have been on an enormous scale to produce the improvement in every single sub-test, shown in Table 3. Also, the effect of failure to observe the rules would have been to reduce the correlation between the tests, and the correlation could not have been as high as it was (Table 5) if this had occurred on a large scale. A much more likely effect of the difference in conditions is a difference of attitude to the test. On the one hand, several graduates commented in letters that they were able to tackle the test on the second occasion without anxiety, and anxiety is known to depress scores. Likewise, some shied away from the forbidding number series sub-test on the first testing, but tackled it competently on the later occasion. Conversely, some may not have made a serious effort on the first occasion. Those who teach psychology students will be familiar with the person whose defence, when presented with a test, is to treat it with levity : in their mature years, such persons may approach a test more conscientiously. Experience suggests that the number of students who decline to participate seriously in testing is small ; and though some of the very large increases in score may have resulted from this, even the omission of all those with increases of over twenty points of score would still leave a highly significant average improvement in score.

The second spurious factor is test-sophistication. A group in which the majority are teachers is bound to have acquired a familiarity with the content of intelligence tests and the best method of tackling them in the course of the last twenty years. It is most unlikely that any of them had encountered in the interval the actual test used here, since it is too difficult for most children ; but work with similar kinds of tests may have raised their scoring power artificially. Such sophistication would presumably show itself by increased speed on the test as well as by improved performance. As was stated above, internal evidence shows that the second was done at higher speed. However, improved speed was not the only cause of the increase in score, for there was a consistent improvement in the first few sub-tests which were attempted by everyone, as well as improvement in sub-tests near the end. Among the early sub-tests, instructions and number series show the largest increases, and it might be argued that these items particularly favour the person who is already familiar with them. Yet significant increases occurred also in the other early tests which were quite straightforward in content.

Probably both of these spurious influences contributed something to the over-all improvement in score, but the whole of the increase can hardly be explained away in such terms. It would, therefore, seem reasonable to conclude that the improvement in score with age represents, to some extent, a genuine improvement in ability.

Two further points may help to show that this result, confirmed as it is by other work, is not wholly inconsistent with the evidence of a decline of intelligence with age. Firstly, in the six follow-up studies listed, all of which have shown some improvement with age, a considerable amount of the test material in every case has been verbal (though the improvement has not been entirely confined to the verbal items) and verbal abilities are known to be much less affected by age than other types of ability less commonly exercised in every-day life. Secondly, in all six studies, the subjects were of superior ability, had continued their education to an advanced level, and generally were engaged in work which, if not intellectually stimulating, was certainly more exacting mentally than that of the majority of adults. How far this affects test performance is difficult to gauge. There was no significant difference between the amount of increase among honours graduates and among ordinary graduates; nor between the improvement shown by secondary school teachers and by primary school teachers; nor, among the women, between the improvement in score of those in jobs and that of housewives. It can, however, be concluded that the results from this and other work show no evidence of any decline with age in test performance among subjects of advanced education and superior ability up to about the age of fifty.

V.—REFERENCES.

BAYLEY, N., and ODEN, M. H. (1955). The maintenance of intellectual ability in gifted adults. *J. Gerontology*, **10**, 91-107.

BENTZ, V. J. (1953). A test-retest experiment on the relationship between age and mental ability. *American Psychologist*, **8**, 319-20 (abstract).

JONES, H. E. (1955). The analysis of age decrements in mental performance (in preparation). Summarized in *Rev. Educ. Res.*, **25**, 417.

NISBET, J. D. (1954). A twenty-year follow-up of intelligence test scores. *Advancement of Science*, **41**, 13-6.

OWENS, W. A. (1953). Age and mental abilities. *Genet. Psychol. Monog.*, **48**, 3-54.

SWANSON, E. O. (1952). The relation of vocabulary test-retest gains to amount of college attendance after a 24-year period. *American Psychologist*, **7**, 368 (abstract).

WECHSLER, D. (1944). *The Measurement of Adult Intelligence* (3rd edition). Baltimore: Williams and Wilkins.

(*Manuscript received* 5*th July*, 1956.)

V.—AN EXPERIMENT ON PRACTICE, COACHING AND DISCUSSION OF ERRORS IN MENTAL TESTING.

BY A. W. HEIM AND K. P. WATTS
(*Psychological Laboratory, University of Cambridge*).

I.—*Introduction.* II.—*Design of experiment.* III.—*Results.* IV.—*Discussion.* V.—*Appendix.* VI.—*References.*

SUMMARY.—An experiment was conducted in an attempt to reconcile the conflicting results reported on coaching and practice effects in mental testing. Parallel and identical (spatial) tests were administered weekly, to 13- and 14-year-old boys and girls, drawn from six schools. On the basis of first test score, every form was divided into five equated sub-groups each of which subsequently received different treatment. Two sub-groups took the *same* test (Test 1) for six consecutive weeks, one of these receiving coaching but not the other. The remaining three sub-groups took a *different* (parallel) test each week, receiving between sessions either coaching or discussion of previous week's errors or neither. At the final session, every sub-group took Test 7.

The experimental results are confused by the fact that inter-school differences equalled or exceeded inter-group differences. The conclusions which follow are, therefore, offered tentatively: (*a*) Discussion of errors is the most effective improver of performance on a *new* test. (*b*) Coaching plus practice is more efficacious than practice alone, whether the practice task be the re-taking of identical or of parallel tests. (*c*) Coaching plus practice-on-the-same-test produces better performance on a new test than does practice alone on different tests. (*d*) Practice alone on the same test yields the poorest result, on the criterion of performance on a new test. Personality differences among individuals and groups are found to contribute materially to the results.

I.—INTRODUCTION.

A GOOD deal of work has been done in recent years on the relative importance of practice and of coaching on intelligence test performance. Among the many papers on this topic may be cited those by Peel (1952, 1953), Wiseman and Wrigley (1953) and Donaldson (1954). In 1953-4, this *Journal* published a Symposium on the Effects of Coaching and Practice in Tests. As Vernon (1954) indicates in his Conclusions, there is some measure of agreement among the symposiasts, particularly on such points as the association of test-familiarity with some degree of improved performance and " the great individual differences in coach-ability." He discusses also the discrepancies found among the results of different investigators. These he attributes largely to " methodological problems: the large Standard Error (or unreliability) of coaching gains, the expression of gains in non-comparable units, and the failure to control the difficulties of tests that are being compared."

The present writers were struck by the conflicting results and by Vernon's analysis of their causes. They felt, too, that an additional cause might lie in the ambiguity of some of the key terms. For example, the contradictory findings that coaching has more effect than practice (James 1953, Dempster 1954) and that practice has more effect than coaching (Yates 1953, Wiseman 1954) might prove reconcilable if the meanings of ' coaching ' and ' practice ' were clarified at the outset. For many readers, coaching *without* practice is not coaching at all.

This enquiry was initiated in an attempt to overcome some of these difficulties and, thereby, to resolve some of the conflicts. The experimenters aimed in particular to differentiate clearly between (*a*) practice; (*b*) coaching; and (*c*) discussion of errors; to combine the technique of giving the same test repeatedly with that of giving parallel tests; and to ensure that the latter were, in fact, strictly parallel.

The tests used were seven tests of Spatial Perception, known as the Shapes Analysis series (Cane and Horn, 1951). Their level of difficulty had previously been equated, on the dual criteria of proportion of subjects answering correctly and of time taken per question. A description of these tests is given in the Appendix. The Shapes Analysis tests were chosen partly because they were believed to be strictly equated for level of difficulty and homogeneity and partly because the vast majority of previous researchers on this topic have used intelligence tests exclusively. In addition to the seven spatial tests, however, an intelligence test, AH 4 or AH 5 (Heim, 1955, 1956) was given at the beginning and again at the end of the experiment in an endeavour to estimate transfer effects. Differences between the performance of boys and girls, and of two age groups were also to be examined.

The subjects were drawn from six schools: two secondary modern girls' (P and R), two secondary modern boys' (Q and S), one central girls' (T) and one boys' grammar school (U). Table 1 shows the arrangement of test and control groups according to school and age. It may be seen that six test and six control groups were used, of whom half the subjects were 13-and half were 14-year-olds. In general, the subjects were drawn from A-stream forms; the only exceptions were Test group U and Control group T, who were B-stream children.

AH 5, a test of high grade intelligence, was given to the 14-year-olds and to the (13-year) grammar school boys; AH 4, an easier intelligence test, was given to all the other 13-year-olds. Both these intelligence tests are in two parts, Part I having a verbal and numerical bias, Part II having a diagrammatic bias.

TABLE 1

EXPERIMENTAL DESIGN.

	SECONDARY MODERN SCHOOLS					CENTRAL SCHOOL			GRAMMAR SCHOOL		
	P Girls	Q Boys	R Girls	S Boys	Intelligence Test	T Girls		Intelligence Test	U Boys		Intel-ligenc Test
13-year-olds	Test Group N=29	Test Group N=35	Control Group N=40	Control Group N=33	AH 4				Test Group N=22	Control Group N=31	AH
14-year-olds	Control Group N=33	Control Group N=35	Test Group N=35	Test Group N=34	AH 5	Test Group N=27	Control Group N=24	AH 5			

II.—DESIGN OF EXPERIMENT.

Table 2 shows the test programme for any one set of subjects. Each session consists of a double school period, and the interval between testing sessions is one week. Thus, in the Test group, each form at the preliminary session takes the intelligence test. A week later, in Session I, they all take Shapes Analysis (S.A.) Test No. 1, preceded by illustrative Examples. On the basis of their S.A. results, each form is divided into five sub-groups (A, B, C, D and E), equated for score, range and number of questions attempted. Intelligence test scores also are taken into account, in the equating of sub-groups.

From Session II onwards, the five sub-groups receive different treatment. A and C have a normal school lesson for the first half of each session (except for the grammar school boys who have a free period). During this time, one of the experimenters coaches B and D, while the other experimenter gives E ' discussion of errors ' on their previous week's performance. During the second half of each session, A and B weekly take S.A. Test No. 1, making six times in all. C, D and E take a different S.A. test every week. Preliminary examples are dispensed with after the first S.A. testing. This procedure is followed up to and including Session VI. At Session VII, A and B thankfully abandon Test No. 1 and all five sub-groups take S.A. Test No. 7. At the final session (ninth week of experiment), all subjects take the intelligence test for the second time.

Every form in the control group takes the intelligence test at the preliminary session and again, eight weeks later, at the final session. They do no S.A. test at any time.

In this paper, ' coaching ' signifies working through with the subjects problems which closely resemble, but are not identical with, the problems comprising the tests. The children are encouraged to take an active part: to offer suggestions, to ask questions and to explain solutions. Thus ' coaching ' here includes some practice; and, of course, all the coached children along with the the others had straightforward practice, during the second period of each session. ' Discussion of errors ' consisted in going over with the children about a dozen questions which they had answered wrongly the preceding week. The investigators made a point of choosing questions on which the majority of the children had failed. As a morale-builder, however, they ensured that every child in the group had, in fact, answered at least one of the crucial questions correctly. Each child was given an answer sheet, on which were recorded his answers to the ten or twelve questions selected for this purpose. The remaining spaces on these prepared answer sheets were left blank. Individual total scores were not revealed.

TABLE 2

TESTING PROGRAMME.

		Procedure during first part of session.	Preliminary Session	Sessions :							Final Session
				I	II	III	IV	V	VI	VII	
				Shapes Analysis Test No. :							
Test Group	Sub-group A	Free period or ordinary lesson ..	Intelligence test	1	1	1	1	1	1	7	Intelligence test
	Sub-group B	Coaching	„	1	1	1	1	1	1	7	„
	Sub-group C	Free period or ordinary lesson ..	„	1	2	3	4	5	6	7	„
	Sub-group D	Coaching	„	1	2	3	4	5	6	7	„
	Sub-group E	Discussion of Errors in previous test ..	„	1	2	3	4	5	6	7	„
Control Group		....	„	ordinary school lessons							„

The figures constituting the seven tests were fairly small, three questions per quarto page. The figures for ' coaching ' and for ' discussion of errors ' were first presented in the same form and were discussed until general agreement was reached among the children as to the right answers. Only then were the ' solution ' figures demonstrated. These figures were drawn specially for this purpose : they were larger than the original test questions (one drawing per page) and were brightly coloured. It was hoped that the colours might increase interest and also prove useful for illustrative purposes (e.g., if the answer to a ' how many units ' question was nine, there might be three red, three yellow and three blue units in the illustration).

The two experimenters adopted as far as possible similar methods and manner. In order to minimise any effects attributable to personality factors, however, they alternated weekly between the coaching and discussion of errors groups. No individual member of any group was told his score until the end of the experiment when the few outstanding scorers received a small prize ; no other individual results were ever made public. After the fourth week of testing the mean score of each sub-group was announced at the beginning of the test-period. This was done in order to maintain interest. In several schools a definite team spirit was established, the announcement of the scores being greeted with enthusiastic cheers or boos by the appropriate sub-groups.

III.—RESULTS.

A.—Selection of subjects for analysis.—When the data had been collected, the first problem was to decide which subjects, if any, to exclude from the analysis of the results. Since the experiment extended over nine consecutive weeks of the school term some of the children, inevitably, were away for several of the testing sessions. We were reluctant to exclude from the analysis any subjects who had taken both the intelligence test and the first Shapes Analysis test, as their scores had been used in equating the sub-groups. On the other hand, to include a child who had missed three or four sessions or who was absent for the all-important seventh S.A. testing seemed equally unsatisfactory. To include only those subjects who took every single test would have sadly depleted the numbers in the sub-groups. For instance, in one set (U), sub-group D would have been reduced to two subjects—who happened to be the lowest two scorers in the sub-group, with an initial mean of 5·5, the mean for the whole sub-group being 10.0.

We, therefore, compromised. The analysis which follows is based on the results of the (182) children who took the intelligence test both times, the first and last S.A. test, and at least three of the intermediate S.A. tests.

B.—Progress curves of sub-groups.—The clearest overall picture of the results is probably given by the mean progress curve of each of the five sub-groups. In Figure 1, these curves are shown for all the subjects combined, regardless of age, sex and school. The continuous line shows the mean for all these children. This latter curve is difficult to interpret, since two-fifths of the subjects (sub-groups A and B) took the same test for the first six sessions whereas the remaining three-fifths (C, D and E) were taking a different test each week. It does, however, show a continuous rise up to the sixth session and a fall from Sessions VI to VII. It may be seen that the biggest contributors to this fall are sub-groups A and B, who at Session VII were taking a new test for the first time since the beginning of the experiment. But even D and E. show a slight decline on the seventh testing. Sub-group C (different tests, no coaching or discussion of errors) appears to have reached a low level plateau by the fourth or fifth session.

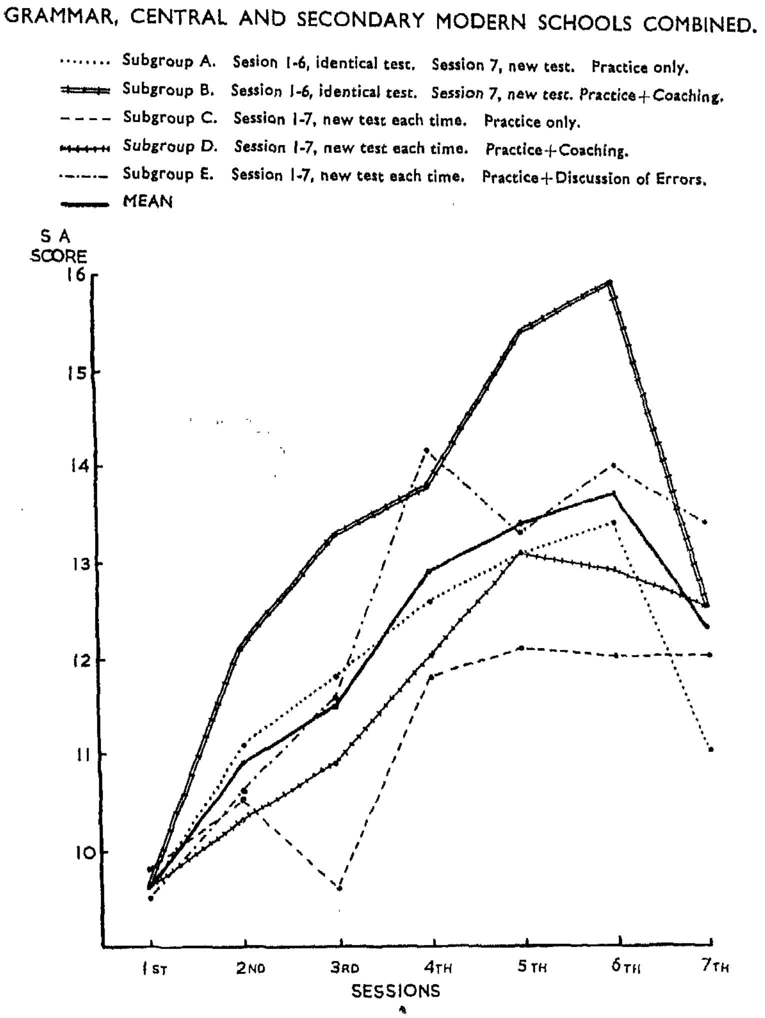


FIG. 1

Progress Curves of Subgroups

Despite the general tendency to improve with retesting, not every sub-group reached its peak as late as the sixth session. In fact, this occurred only with A and B—the two sub-groups who took the same test for six consecutive weeks. The highest *final* score was achieved by E, the discussion of errors sub-group. But their own peak was reached in Session IV and this was considerably lower than B's peak, attained on sixth taking of Test 1. This was the highest score gained at any time by any sub-group.

It is interesting to compare B's performance with that of A, who followed the same programme apart from lack of coaching. B drew ahead of A from the very beginning; the difference between their mean scores steadily increased, with the exception of Sessions IV and VII. If we now compare C and D, a somewhat similar picture is found. It will be remembered that both of these sub-groups took a different test each week. But C, like A, had practice only; and D, like B, had coaching in addition. In general, D did better than C but the difference between these two sub-groups, throughout the testings and at the end, is smaller and less consistent. The differential effects of practice and coaching appear to be less on different tests than on identical ones. There is no difference on seventh testing between sub-groups B and D.

E, the sub-group which had discussion of errors, has done best at the seventh session—when everybody was taking a new test—but they were eclipsed by B (same test, coaching) in most of the earlier testings. Thus, if we confine ourselves to performance on a *new* test (Session VII):

(*a*) Discussion of errors would appear to be the most effective method of producing improvement.

(*b*) If coaching plus practice is given, it appears to matter little whether the practice is on identical or on similar tests.

(*c*) Coaching plus practice-on-the-same-test is more efficacious than practice alone on different tests, but

(*d*) Practice alone on different tests does better than practice alone on the *same* test, the latter procedure yielding the poorest results of all.

In order to check these findings, progress curves were drawn also for those (139) subjects who did not miss a single test session; and also for the scores expressed in terms of proportion of correct to attempted. The relationship between the five progress curves remained unchanged, with one exception: on ratio of number correct to number attempted, B did better than D at the seventh session.

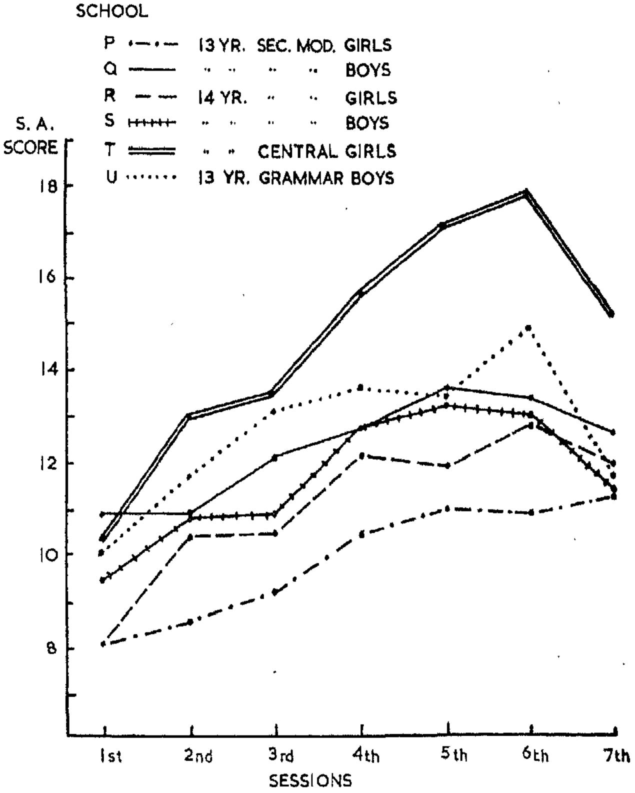
In order to estimate the significance of these results, a two way analysis of variance (between schools and between sub-groups) was performed. Absolute score would have been an unsatisfactory criterion owing to the variation among schools on initial test score (see Figure 2). Increase in score was, therefore, used. The only difference between pairs of sub-groups at seventh testing found to be significant at the 0·05 level, was that between E and A. On sixth testing, B was significantly higher (at the 0·05 level) than both C and D, none of the other pairs at this testing differing significantly from one another.

The reasons for these somewhat negative results appear to be two-fold. (*a*) The differences between schools were as great as, or greater than, the differences between sub-groups. (*b*) Score increases were very spasmodic—more so even than would appear from Figure 1 in which the adding of the six schools has smoothed out some of the wilder waves.

Fig. 2 illustrates the first of these points. In it the five sub-groups are combined, enabling a general comparison of school with school to be made. The range on initial score may be seen here and the (somewhat surprising) rank order of the six schools. This figure shows also that reasonably high consistency was maintained, as between schools.

To give the curves for the separate sub-groups school by school would be too space-consuming. These curves show a good deal of criss-crossing, the gains and drops from week to week being extremely erratic. The progress curves of individual subjects naturally tend to be even less consistent.

C.—Sex differences and test score.—The chaos of the total results is reducible to some semblance of order if the major groups are divided into smaller units. This holds whether the principle of classification be age or sex or form or school, i.e., on any one of these divisions there appear systematic-looking results susceptible of reasonable interpretation. The pity is that each set of results so obtained produces a *different* and conflicting set of findings and interpretations. It seems scarcely worthwhile, therefore, to present the many classifications which are available given this experimental design. On the other hand, one such illustration may be desirable.



SUBGROUPS A TO E COMBINED

FIG. 2
MEAN PROGRESS CURVES OF SCHOOLS

The choice is made of Figure 3 on the joint grounds of intrinsic interest and of providing a striking example of the above phenomenon. In order to compare with one another groups of broadly similar intelligence and educational level, only secondary modern school children are included. The two main comparisons are (*a*) between girls and boys (combining both age groups) and (*b*) between 13- and 14-year-olds (combining both sexes).

If, then, the top half of the figure be considered, it becomes apparent that *with girls* ' identical retesting plus coaching ' (sub-group B) and ' discussion of errors ' (sub-group E) are far more efficacious than any of the other three methods; *with boys*, however, it is evidently ' discussion of errors ' alone which clearly differentiates after three testings. If we now inspect the lower half, we find that *13-year-olds* in general have something in common with

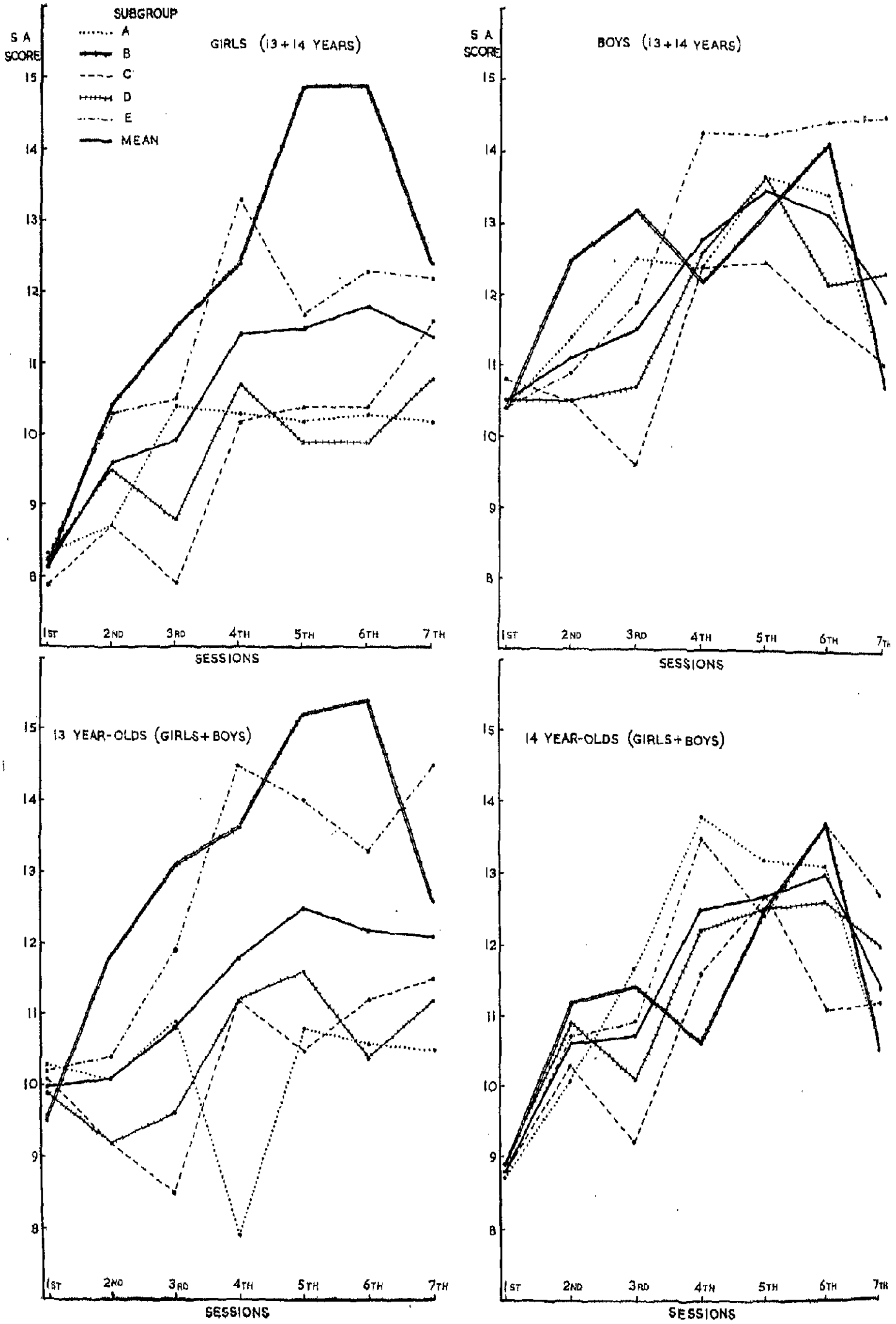


FIG. 3

Secondary Modern Schools

the *girls* above : they too do best with ' identical retesting plus coaching ' and ' discussion of errors.' On the other hand, for best score on a *new* test, the efficacy of these two methods is not constant : observe the changed order of the top pair of curves in the two left-hand figures. With *14-year-olds*, it may be seen that the five methods in general produce remarkably similar results. It is hardly necessary to add that the remaining two schools, grammar boys and central girls, present a different picture again from any of those in Figure 3. An overall comparison between the 13-year-olds and 14-year-olds reveals a wider range among the younger children and a *higher* initial and final mean score.

A comparison between the sexes, irrespective of age, confirms the findings of previous investigators using spatial tests, the boys scoring more highly initially, but effecting slightly less improvement, than the girls. A two-way analysis of variance (between sexes and between sub-groups) for 'improvement scores' showed that whether improvement between first two testings be considered (second score minus first score) or improvement on the experiment as a whole (seventh score minus first score), the difference between the boys' and girls' rates of improvement is statistically significance at the 0·05 level.

There is one further difference in test performance between the boys and girls, which may be of interest, namely, the differences in the intelligence tests—S.A. correlations. These are shown in Table 3.

TABLE 3

INTER-CORRELATIONS BETWEEN S.A. AND INTELLIGENCE TESTS.*

		Boys	Girls
Intelligence Test :		(62)	(53)
AH 4	Part I	0·17	0·49
	Part II	0·25	0·59
	Total Score	0·27	0·61
		(91)	(95)
AH 5	Part I	0·13	0·33
	Part II	0·53	0·57
	Total Score	0·45	0·56

* These figures include all the children who at any time took an S.A. test and AH4 or 5. Score on first testing only is used for these correlations.

The second parts of the intelligence tests would be expected to correlate more highly than the first parts with the S.A. tests since the second parts are diagrammatic, whilst the first parts are verbal and numerical. But it is surprising that in AH 4, total score correlates more highly with S.A. than does Part II. Of greater interest perhaps is the fact that the girls' intercorrelations should be consistently higher than the boys', significantly so in the case of AH 4, Part II and total scores (0·05 level of confidence). This suggests that those girls who do well on the spatial test may do so rather by virtue of high general intelligence than of marked spatial aptitude.

D.—Miscellaneous.—This section constitutes a sort of waste paper basket into which has been thrown some of the problems which we had hoped to illuminate but whose obscurity has, if anything, been deepened by the experimental results. We had aspired to find an answer to such questions as the following : What effect does training on a spatial test have on intelligence test performance ? When various methods of training are used, do the initially higher or lower scorers benefit more ? Does the particular method used have a bearing on this problem ? Is it the case that practice and coaching on mental tests effect improvement only during the first few sessions—as reported, for instance, by Bernstein (1924) ? Or does the improvement continue as long as the opportunity is open—as suggested, for instance by Heim and Wallace (1950) ?

The data already presented suggest that conclusions on these latter questions are largely speculative since the various progress curves are so discrepant and are, individually, so unpredictable. If we consider, for example, the mean in Figure 1, a general impression is gained that whilst the gradient was becoming progressively gentler, the mean scores continued to rise, up to and including Session VI. At the seventh session, however, a tendency for scores to drop is observable in sub-groups D and E (new test each week) as well as in A and B (new test for first time at Session VII). Moreover inspection of the separate sub-groups suggests that C and D respectively, had reached a plateau by Sessions IV and V. Thus, no conclusions of any value can be drawn from the experiment on the problem of duration of improvement in test performance.

Similarly inconclusive and negative are the results concerning transfer effects of spatial tests on intelligence tests. As was to be expected, all groups tended to score more highly on second than on first intelligence testing. The question was whether the *test groups* (who had taken the S.A. series) would show significantly more, or perhaps less, improvement than the *control groups* (who had no test-training of any kind between their two intelligence test sessions). When, however, the comparison was made between the increase in intelligence test score shown by the test groups and the control groups, the differences were found to be not only small and statistically non-significant, but also such differences as there were did not even go in a uniform direction.

IV.—Discussion.

Despite the inconclusiveness of the experimental results, it is felt that this investigation is not entirely valueless. The experiment was initiated largely in an endeavour to reconcile or explain the many conflicting results reported to date, the association between number of investigators and number of conclusions having proved even closer than is customary in psychological research. Far from clarifying the situation, however, we have merely achieved *internally* conflicting results.

It is possible, of course, to explain away some of the anomalies with reference to the ' personalities ' of the forms and the sub-groups. For example, some forms (notably our T school girls) were exceptionally keen and lively whilst others (in particular, our U school boys) were uninterested, and exhibited the passive resistance at which schoolboys excel when so inclined. These impressions were confirmed by the school staffs who had, however, given us no advance information along these lines. It is natural that such differences exist and that they should interact with the different treatments accorded to our sub-groups : for instance, an idle bored collection of children is liable to benefit disproportionately little from discussion of errors or from coaching.

Again, the sub-groups, which were equated by us and allocated before we became personally acquainted with any of the children, rapidly evolved their own character and team spirit (or lack of it). This varied greatly from sub-group to sub-group ; it appeared to depend largely on one or two sterling leaders or forceful spivs, as the case might be ; and it undoubtedly played an important part in determining the test performance of the sub-group.

For example (14-year-old) sub-group B, in School S, were a well-grown lot of boys with mature out of school interests—liable to be imported into school life at the least provocation. It may be recalled that sub-group B (and also D) received coaching. It so happened that the room in which their coaching session was conducted had big windows looking out on the playing fields, where members of the corresponding girls' school disported themselves at this time. Inevitably, subjects in B who were moved away from the windows tended to sulk; and subjects remaining by the windows tended to distribute their attention with fine impartiality, between the rectilinear shapes on their desks and the curvilinear shapes on the playing fields. Sub-group D unashamedly kept its nose to the grindstone.

On the other hand, sub-group E, in School T, was the one set of girls in an otherwise responsive form who were bored and unco-operative, and failed to interest themselves in either the test problems or the weekly results. They were, in fact, the only 'leaderless' sub-group of the form. The other four sub-groups were inspired in each case by one or two keen individuals who stimulated interest and fostered friendly rivalry. In all the other schools, the E's responded visibly to the social and intellectual stimulus of the 'discussion of errors' situation. The T school E's, however, were lackadaisical individually and unintegrated as a group. The results of this in terms of test performance were clearly seen, School T leading the field in every sub-group but E.

One further illustration of the effects of boredom in a sub-group: the A's in School P. This school gave in general the worst performance. Sub-group A is interesting, however, for psychological reasons. It was obvious even before looking at their test scores that these girls became tired, and somewhat resentful, at doing the same test over and over again. They intimated that even a little coaching such as their fellow-sufferers in B enjoyed would relieve the tedium. The A's in this school happened to include the top scorer on AH 4 and she radiated a lively distaste for the monotony of Test 1, which communicated itself to the other members of the sub-group A. Whilst every other A group dropped at the seventh session, when presented for the first time with a new test, school P *rose*, managing to exceed school U who had scored far more highly all along.

These social phenomena have been described at some length because, apart from their intrinsic interest, it is felt that they go some way towards explaining the inconsistency of the results. Moreover, it is possible that previous investigators may have ignored or minimised such effects and that they contribute substantially to the widespread disagreement in the field. It had been hoped in this enquiry that the experimental design would take care of such differences, and it is disappointing—though not without precedent—to find that the psychological effects have over-ridden the psychometric effects. Such coherence as does emerge, however, indicates that the different methods probably do produce different results and that their relative efficacy accords with common sense expectations.

ACKNOWLEDGEMENTS.—We wish to express our thanks to the relevant county and city education officers for permission to carry out the investigation; to the head masters and head mistresses of the schools for their co-operation; to Mr. F. Anscombe and Mr. E. G. Chambers for statistical advice and assistance; to Mr. J. W. Whitfield for his help in designing the experiment; and to the Medical Research Council for financial support.

V.—APPENDIX.

SHAPES ANALYSIS TEST SERIES.

The tests were seven chosen from the series known as Shapes Analysis (Cane and Horn, 1951). Each S.A. test consists of 36 problems of which 18 are two-dimensional, and 18 three-dimensional. The problems are of six types :

(i) How many units like the small figure are needed to make the large figure ? (Creative answer. Two-dimensional).
(ii) Which of the five lower figures will remain if the top left figure is taken from the top right figure ? (Choice of 1 from 5 offered solutions. Three-dimensional).
(iii) Which one of the five lower figures cannot be made from a number of units like the small figure above them ? (Choice of 1 from 5 offered solutions. Two-dimensional.)
(iv) Same form of question as (i) but drawings three-dimensional.
(v) Same form of question as (ii) but drawings two-dimensional.
(vi) Which two of the eight lower figures are needed to make the figure above them ? (Choice from 28 possible pairs. Three-dimensional.)

Thus each S.A. test contains six types of question or principle, each principle being represented six times. The types of question occur cyclically throughout the test, in the order given above. The questions are arranged in order of difficulty (within principles) from easiest to hardest but the gradient is not steep, every problem being of a high level of difficulty.

Subjects record their answers on a separate sheet. They are instructed not to mark the question book in any way, but are told that they may use their answer sheet for rough work, should they so desire. A small minority avail themselves of this permission. When taking the test for the first time, subjects are presented with six pairs of preliminary Examples, illustrating the six principles embodied in the test. The solution to the first member of each pair is given on the answer sheet. The second problem of each pair is to be solved by the subject himself, working without a time limit, and receiving a minimum of help from the tester. No one embarks on the test proper until he has completed all the preliminary Examples correctly, and to his own satisfaction.

A time limit of 30 minutes is imposed for the test, exclusive of these Examples. On first testing, the subjects attempted an average of 27 out of the 36 questions. One mark is awarded for each correct answer ; neither omissions nor errors are penalised.

VI.—REFERENCES.

BERNSTEIN, E. (1924). Quickness and intelligence, an enquiry concerning the existence of a general speed factor. *Brit. J. Psychol. Monogr., No. 7.*

CANE, V. R., and HORN, V. (1951). The timing of responses to spatial perception questions. *Quart. J. Exp. Psychol.*, **3**, 133-145.

DEMPSTER, J. J. B. (1954). Southampton investigation and procedure. *Brit. J. Educ. Psychol.*, **24**, 1-4.

DONALDSON, M. C. (1954). Effects of coaching and practice on intelligence test scores. Abstract of paper. *Bull. Brit. Psychol. Soc.*, No. 23.

HEIM, A. W., and WALLACE, J. G. (1950). The effects of repeatedly retesting the same group on the same intelligence test : Part II. High grade mental defectives. *Quart. J. Exp. Psychol.*, **2**, 19-32.

HEIM, A. W. (1955). *Manual for the Group Test of Intelligence AH* 4. National Foundation for Educational Research.

HEIM, A. W. (1956). *Manual for the Group Test of High Grade Intelligence AH*5. National Foundation for Educational Research.

JAMES, W. S. (1953). Coaching for all recommended. *Brit. J. Educ. Psychol.*, **23**, 155-162.

PEEL, E. A. (1952). Practice effects between three consecutive tests of intelligence. *Brit. J. Educ. Psychol.*, **22**, 196-199.

PEEL, E. A. (1953). Footnote on " Practice effects." *Brit. J. Educ. Psychol.*, **23**, 126.

VERNON, P. E. (1954). Conclusions on symposium. *Brit. J. Educ. Psychol.*, **24**, 57-63.

WISEMAN, S., and WRIGLEY, J. (1953). The comparative effects of coaching and practice on the results of verbal intelligence tests. *Brit. J. Psychol.*, **44**, 83-94.

WISEMAN, S. (1954). The Manchester experiment. *Brit. J. Educ. Psychol.*, **24**, 5-8.

YATES, A. (1953). An analysis of some recent investigations. *Brit. J. Educ. Psychol.*, **23**, 147-154.

(*Manuscript received 7th August*, 1956.)

VARIABILITY IN ATTAINMENT

By LESLIE C. D. KEMP

(*Armidale Teachers' College and University of New England, Armidale, N.S.W., Australia*).

I.—*Introduction.* II.—*Method and results.* III.—*Conclusions.*

SUMMARY. 1.—Growing out of an earlier investigation into certain factors associated with level of attainment in junior schools, a subsidiary study has been made of the variability in attainment within schools.

2.—Highly significant differences in spread of attainment scores within schools have been found.

3.—Lack of interest in school appeared to be the only characteristic (of those considered) associated with greater variability in attainment in all the school subjects tested.

4.—Variability in attainment in what have been called the *rote* subjects appeared to be greater in schools where, in addition to a lack of interest, there was a lower level of intelligence, socio-economic status, and morale, with a progressive rather than traditional type of organisation. These schools tended to be housed in older types of buildings with larger playground space per pupil.

5.—Less definite indications were found that variability in the *comprehension* subjects tended to be greater in schools where the children, in addition to showing lack of interest, came from homes of higher socio-economic status. These schools tended to be large.

6.—There is a suggestion that a more progressive kind of school organisation permits greater variability in rote attainment, while the more traditional or formal kind of organisation tends to lead to more homogeneous achievement.

7.—Size of class appears not to be associated with variability in attainment.

I.—INTRODUCTION.

AN earlier article in this *Journal** reported the results of an investigation into the relationships of various characteristics of a number of junior schools and their environments to one another and to the average level of attainment in the final junior year. The characteristics, which were assessed by various techniques, included the school site and buildings, the homes and neighbourhood of the school, the attitude of the head, and the staff, the behaviour of children, and the type of school organisation. It was found that the main factors determining level of attainment in the formal school subjects were intelligence level, socio-economic status, size of school, and morale. Morale was found to be determined not only by the attitudes of the head and the staff, but by the intelligence level and socio-economic status of the school. Interest in school, though not significantly correlated with morale, was found to be also partly determined by intelligence level and socio-economic status, and to be stronger in voluntary than in county schools. Progressiveness of school organisation (i.e., a more informal and permissive school atmosphere, with a curriculum organised through activities and experiences related to the needs and interests of the children) was found to be very slightly related to attainment in the formal school subjects, to morale, and to interest in school.

During this earlier investigation it was noticed that there were differences in variability or spread of attainment within the schools used. It was, therefore, thought worthwhile to pursue this matter further, and to endeavour to ascertain what relationships existed, if any, between the various environmental and other factors and variability in attainment.

---

* KEMP, L. C. D. (1955). "Environmental and other characteristics determing attainment in primary schools." *Brit. J. Educ. Psychol.*, XXV, 67-77.

II.—METHOD AND RESULTS.

The school subjects in which attainment was measured were divided on the basis of a factor analysis into two groups designated ' comprehension ' and ' rote.' The former included arithmetical problems, reading comprehension, and general knowledge ; the latter, mechanical arithmetic, spelling, writing and composition. Factor scores were calculated for every child by adding the standard scores in the appropriate tests ; and for every school the variances of the two distributions of factor scores were calculated. For each factor, a simple analysis of variance was carried out to examine the null hypothesis that there were no greater differences between schools in factor-score variance than could be accounted for by sampling errors. Table 1 shows the results for the comprehension factor. It is seen that the F-ratio is large enough to justify the rejection of the null hypothesis with a very high degree of confidence.

TABLE 1

ANALYSIS OF VARIANCE OF VARIABILITY IN ATTAINMENT IN THE COMPREHENSION FACTOR.

Source of variation	Sum of Sqq.	df	Mean square
Between schools	1935·5	49	39·50
Within schools	8772·1	1361	6·44
Total	10707·6	1410	—

F=6·13. (At 5 per cent. level F is less than 1·36 ; at 1 per cent., 1·54).

The results for the rote factor are shown in Table 2, from which once again it can be seen that the null hypothesis that there is no difference in variability of factor score from school to school can be rejected at the 1 per cent. level of confidence.

TABLE 2

ANALYSIS OF VARIANCE OF VARIABILITY IN ATTAINMENT IN THE ROTE FACTOR.

Source of variation	Sum of Sqq.	df	Mean square
Between schools	4526·4	49	92·37
Within schools	14949·4	1361	10·98
Total	19475·8	1410	—

F=8·41.

The clear indication that variability in factor-scores cannot be adequately accounted for on the basis of chance or random errors justifies an investigation into factors associated with this variability. Coefficients of correlation were calculated between the variance within schools in each attainment factor and each of thirteen school characteristics. These are shown in Table 3.

TABLE 3

COEFFICIENTS OF CORRELATION BETWEEN SEVERAL SCHOOL CHARACTERISTICS AND EACH OF TWO ATTAINMENT FACTORS.

Characteristic*	Comprehension factor	Rote factor
Socio-economic status	·24	– ·24
Intelligence level	·08	– ·28
Comprehension attainment	·18	– ·27
Rote attainment	·12	– ·29
Interest in school	– ·25	– ·39
Morale	– ·04	– ·25
Enrolment	·24	·16
Progressiveness	·06	·21
Playground space per pupil	·10	·25
School building	·01	– ·27
Staff spirit	– ·04	·05
Class size	·05	·11
Adjustment	·04	·04

* Further description of these characteristics, see Kemp, *op. cit.*

In this particular investigation a coefficient of correlation must be as large as ·35 to be significant at the 1 per cent. level ; and ·27, at the 5 per cent. level.

III.—CONCLUSIONS.

A marked difference is immediately observable in these two sets of coefficients. Variability in attainment within schools in the comprehension factor is not significantly related to any of the school characteristics, although the correlations with three of them approach the 5 per cent. level of significance. On the other hand, variability in the rote factor is significantly related to five of the characteristics and almost significantly related to four others.

It might justifiably be concluded, therefore, in relation to the comprehension factor, that apart from a suggestion for greater variability to be found in larger schools (·24), in areas of higher socio-economic status (·24), and where the children are less interested in school (– ·25), little light has been shed on the determination of variability in this kind of attainment. In connection with attainment in the rote factor, it seems justifiable to conclude that greater variability tends to be found in schools where the children are less interested in their work (– ·39), are of lower average intelligence (– ·28), and where the average level of attainment is lower (– ·27 and – ·29). These schools tend to be housed in older and less attractive buildings (– ·27). There is a somewhat less definite relationship between greater variability and lower socio-economic status (– ·24), poorer morale (– ·25), greater playground space per pupil (·25), and more progressive kind of organisation (·21).

The only characteristic showing a definite relationship with variability in both kinds of attainment is interest in school. Lack of interest goes with greater variabiility. As was mentioned in the earlier article (*loc. cit.*, p. 74), the schools where interest was greater tended to be schools where the children were more intelligent, and came from homes of higher socio-economic status. They tended to be voluntary schools, and to be housed in older and less attractive buildings. It is now also seen that greater interest is associated with more homogeneous achievement within the school, particularly in the more mechanical subjects.

One possible interpretation of the correlations obtained is that the schools with low variability of attainment are tending to bring their pupils to much the same level of attainment, while high variability probably means that teachers are encouraging the brighter children to forge ahead and are neglecting the duller ones. The correlations suggest, too, that the former is happening more often in schools of the more formal kind of organisation ; while the wider scatter of scores is associated with the freer and more progressive kind of school. This is particulalry noticeable in rote attainment. There is less spread of scores in schools exerting more formal pressure. Greater variability in rote attainment is associated with lower socio-economic status, lower morale, lower intelligence level and weaker interest. Evidently in schools of higher socio-economic status the pressure from both the homes and the school brings about greater uniformity in attainment in the pupils of those schools. But the children's liking for school seems to be the most important factor in reducing their variability.

(*Manuscript received 18th August*, 1956.)

SUMMARIES OF RESEARCHES REPORTED IN DEGREE THESES

The Construction and Evaluation of an Interest Test Battery for Technical Science.

BY ROBERT C. S. HAMILTON.

(*Summary of thesis submitted and approved for the degree of B.Litt., University of Glasgow,* 1956.)

I.—AIMS OF THE INVESTIGATION.

The aims of this study were :

(1) To construct an Interest Test Battery for technical science.

(2) To investigate, using the test battery, the inter-relations between ability, achievement, and interest in technical subjects.

(3) To study the use of the test battery as a means of selection for technical education.

II.—THE TEST BATTERY.

(1) *The Practical Information Test* involved the identification of twenty objects which might be of interest to a mechanically-minded boy. The score was the number identified.

(2) *The Written Information Test* consisted of a set of twenty technical items and twenty non-technical items of equal difficulty, arranged in random order, to be answered as a multiple-choice test. The score was the number of correct answers to technical items relative to the total number of correct answers to all the items.

(3) *The Book-Titles Test* was made up of twenty technically biased book titles and twenty non-technical titles of equivalent popularity ; the latter including such items as sports, etc., as distractors to give sharper discrimination power. These were presented in random order to the subjects, who were asked to record whether or not they would like to have that book, say, as a prize. The score was the number of technical titles favourably selected relative to the total number of titles favoured.

(4) *The Preference Grid* consisted of a set of ten statements of ways of occupying leisure time, with a technical bias, the items being arranged in order of popularity, and a corresponding set of non-technical items. A 10 × 10 design was thus constructed in which each technical item had to be compared with a corresponding non-technical item of less, equal or greater popularity. The score was the number of technical items preferred.

(5) *The Pictorial Test* consisted of a set of 10 sheets of pictures, each sheet containing two pictures with a technical background, and four others with various non-technical backgrounds such as sports, animals, human interest, etc. The technical pictures were awarded a scale-value using Thurstone's method of equal-appearing intervals. Non-technical pictures, of course, had a zero score. The subjects were asked to select the two pictures which interested them most on each sheet. The score was the total of the scale-values for all the technical pictures chosen.

The reliabilities of the five tests are shown in Table 1 :

III.—ADMINISTRATION OF THE BATTERY.

The battery of five tests outlined above was given to a sample of 100 boys following a modified junior secondary course in a Glasgow school. This course is heavily biased in technical subjects. (woodwork, metalwork and technical drawing).

The I.Q. range fo the boys was 70-104 with a mean of 88 and an S.D. of 7. The age range was 12 years 6 months to 14 years 9 months, with a mean of 13½ years and an S.D. of 7 months.

TABLE 1

Test	Reliability " r "	" r " Corrected*	Method Used	Sample N=
Practical Information	0·69	0·82	Split-Half	102
Written Information	0·56	0·72	Split-Half	105
Book-Titles Choice	0·62	0·77	1st half 2nd half	112
Preference Grid	0·50	0·66	Split-Half	67
Pictorial Test	0·75	—	Rest/Re-test	94

* Using the Spearman-Brown formula.

The test battery, the I.Q., and the standardised scores made in the school examinations in English, Science, Art and Technical subjects provided ten variables which formed the basis of the correlation matrix shown in Table 2. All correlations were calculated by the product-moment formula.

TABLE 2

CORRELATION MATRIX BASED ON SAMPLE OF 100 BOYS.

	English	Science	Art	Technical	I.Q.	Practical Information	Written Information	Book-Titles Choice	Preference Grid
Science	·36								
Art....................	·17	·13							
Technical	·23	·31	·38						
I.Q.	·48	·50	·25	·34					
Practical Information	·10	·04	·07	·19	·12				
Written Information	–·14	–·00	·06	·10	–·19	·19			
Book-Titles Choice	–·15	–·10	·11	·08	–·10	·30	·14		
Preference Grid	–·03	·06	·04	·07	·06	·32	·16	·46	
Pictorial Test	–·09	·09	·10	·09	·09	·01	·05	·24	·27

Values of " r " required for significance :
At 5% level........0·197
At 1% level........0·256

IV.—THE FACTOR ANALYSIS.

A factorial analysis was carried out using Thurstone's Centroid Method. Three factors were extracted, the second and third being bipolars. Rotation of axes to " simple structure " was carried out in two planes by graphical methods. The factor loadings after rotation of axes are shown in Table 3. The variance due to each test and the percentage variance due to each factor are also shown in this table.

TABLE 3

CENTROID ANALYSIS. FACTOR LOADINGS AFTER ROTATION OF AXES.

	I^{11}	II^1	III^1	K_1^2	K_2^2	K_3^3	h^2
English	·567	−·148	·073	·320	·022	·005	·347
Science	·647	·040	·015	·417	·002	·000	·420
Art	·276	·008	·437	·076	·000	·190	·266
Technical	·422	·029	·534	·177	·000	·284	·463
I.Q.	·789	·003	·023	·623	·000	·000	·625
Practical Information	·058	·366	·315	·003	·134	·099	·236
Written Information	−·191	·232	·260	·036	·053	·068	·158
Book-Titles Choice	−·137	·598	·231	·020	·358	·053	·429
Preference Grid	·084	·749	·041	·007	·560	·002	·569
Pictorial Test	·104	·340	·010	·011	·116	·000	·127
Percentage Variance				16·9	12·5	7·0	36·4

The three factors extracted from the analysis have been interpreted as :

(i) A general intelligence factor.
(ii) A technical interest factor.
(iii) A performance factor.

The general intelligence factor has the highest variance (this would probably be higher if the group were not so highly selected with respect to I.Q.). The interest factor appears to be nearly twice as influential as the performance factor.

V.—CONCLUSIONS.

The inter-correlations between intelligence, achievement and interest in technical subjects are as follows :

(1) There is a significant correlation between general intelligence (measured in terms of I.Q.) and technical achievement.
(2) There is zero correlation between general intelligence and interest in technical subjects.
(3) The correlation between technical achievement and technical interest is insignificant.

However, it appears that, among the tests in the interest battery, Practical Information provides the best individual indication of possible success in technical subjects.

From an inspection of the factor loadings in Table 3 it would seem that the Preference Test is the best individual indicator of technical interest.

The Effect on Ability in Woodwork of Previous Training in Cardboard Work.

BY J. A. CROFTON.

(*Abstract of thesis submitted for the degree of Master in Education in the University of Manchester,* 1955.)

The investigator sought to discover whether an organised course in cardboard work carried out during the final year in the primary school would assist boys to achieve better results when they embarked on a woodwork course in the secondary school.

THE EXPERIMENT.

A sample of 330 boys of the 10 to 11 age group was drawn from twenty-five Oldham primary schools. Three groups were formed, each being a rough cross section of the age group.

The first group, consisting of 110 boys, carried out a graded scheme of cardboard work, necessitating the use of manilla, thin card, and strawboard of various weights. The second group, which numbered 112 boys, carried out craftwork which included some incidental cardboard work; and the third group, numbering 108 boys, executed craftwork which involved the use of materials other than cardboard. It was arranged that the three groups should work through the agreed courses for one year. During the early part of the year each boy took Moray House Intelligence Test No. 48 and the Devon Interest Test.

The following year, when the boys had been transferred to secondary schools, a year's woodwork course was taken which covered certain basic techniques.

The main criterion of ability in woodwork was a woodwork test devised by the author. It was given in its final form at the completion of the year's woodwork course. The practical test was devised to test the achievements of the boys in skills which are important and frequently used during a first year's woodwork course. Conditions and materials were kept as uniform as possible. Two raters marked the test. A template was used in order to standardise marking.

Teachers' appraisals of performance in woodwork course work were used as an independent criterion; their marks, given at the end of the first term and at the end of the first year, were kept as objective as possible. These scores were based on all-round woodwork ability with special reference to *confidence* indicating how to tackle a job; *accuracy*, which is largely dependent on correct measurements and method; *speed*, which can be measured by time tests; and *handling of tools*, which is right or wrong according to the method taught by the teachers.

THE RESULTS.

An analysis of the intelligence test results showed that the means for the three groups were very similar. There was no significant difference except between the 'incidental cardboard' group and the 'non-cardboard' group.

The Devon Interest Test results indicated that all groups showed a low interest in practical activities, but of the three groups, the 'incidental cardboard' group showed the highest measure of interest, and would on this account be expected to display greater achievement in woodwork than the other two groups, particularly when all groups were very similar in intelligence.

TABLE 1

INTELLIGENCE AND INTEREST SCORES.

Group	Intelligence		Interest	
	Means	S.Ds.	Means	S.Ds.
Incidental Cardboard	55·908	12·450	16·031	9·840
Organised Cardboard	53·684	11·575	11·958	10·566
Non-Cardboard	52·457	10·565	12·641	8·745

Judging by the estimated marks supplied by the handicraft teachers at the end of the first term, the 'organised cardboard' group had taken the lead in woodwork achievement. The 'incidental cardboard' group and the 'non-cardboard' group proved to be close in second and third positions respectively. By the end of the year when the teachers again gave estimated marks, it was evident that the 'organised cardboard' group had increased its lead over the 'incidental cardboard' group, which in turn had moved further ahead of the 'non-cardboard' group.

In the practical tests, there was a reasonably good correlation between the marks of the two raters for all groups. The group means were in the same rank order, but the gaps between the groups had widened.

TABLE 2

FIRST TERM, FIRST YEAR AND PRACTICAL TEST SCORES.

Group	First Term	First Year	Practical Test	Correlations	
				First Term First Year	First Year Practical Test
Organised Cardboard . .	12·600	13·863	12·126	r= ·545	r= ·508
Incidental Cardboard . .	11·990	12·449	9·969	r= ·526	r= ·256
Non-Cardboard	11·413	11·555	7·847	r= ·542	r= ·244

In certain secondary schools boys in all three groups were brought together into the same classes and were taught by the same teachers. It was thought likely that this would result in greater uniformity of training and conditions. The results of these specially placed boys exhibited the superiority of the 'organised cardboard' group over the other two groups; but the 'incidental cardboard' group and the 'non-cardboard' group wavered between second and third positions, in contrast to the results of the main analysis.

TABLE 3

MEANS OF MAIN GROUPS AND MIXED GROUPS WITH THE SAME TEACHER.

	Groups with the same Teacher	Main Groups
	First Term Mean	First Term Mean
Organised Cardbroad Group	12·667	12·600
Incidental Cardbroad Group	11·580	11·990
Non-Cardboard Group	11·395	11·413
	First Year Mean	First Year Mean
Organised Cardboard Group	13·403	13·863
Incidental Cardboard Group	12·060	12·449
Non-Cardboard Group	12·116	11·555
	Practical Test Mean	Practical Test Mean
Organised Cardboard Group	12·491	12·126
Incidental Cardboard Group	10·280	9·969
Non-Cardboard Group	8·187	7·847

CONCLUSIONS.

(1) The experiment indicates that a preparatory course of organised training in cardboard work improves subsequent achievements in woodwork.

(2) There is some suggestion that preparatory work with cardboard, even when used incidentally, is perferable to no preparatory cardboard work at all.

The Attitudes of Training College Students Towards Corporal Punishment.

By Mabel Kissack.

(*Abstract of thesis submitted for the degree of Master in Education in the University of Manchester,* 1956.)

The main purpose of the work was to find whether certain factors affected the strength and direction of students' attitudes towards corporal punishment, and to investigate the possibility of influencing these attitudes in a desired direction by emphasising certain aspects of the training college course.

Method.

An attitude scale, consisting of two parallel forms, was devised for the experiment, according to the method of Thurstone and Chave.

The attitudes of students from three different types of training college were measured at the end of their two-year course of training, and compared with respect to college, sex and the age-group of children they had been trained to teach. These comparisons were made by means of analyses of variance, and tests of significance of differences between means.

The attitudes of fifty-seven women students, beginning their training at a women's training college, were measured, and compared with their intelligence as measured by N.I.I.P. Group Test 33. Their attitudes were re-measured after the first school practice, and compared with respect to the age-group they had taught, and the attitudes to corporal punishment of the schools where they had practised. These attitude scores were compared with their marks in the Practice and in the Principles of Education, and with the ratings of an essay on " Spare the rod and spoil the child."

From the results of the first attitude test a random sample of twelve students was chosen to run the college play centre for fourteen weeks, in addition to taking part in the lectures and discussions on educational topics with the rest of the group. At the end of this fourteen-week period, all the students were re-tested and changes in attitudes were examined.

Fifty women teachers, whose attitudes had been measured at the end of their training, took the attitude scale after teaching full-time for 10 months.

Results.

The Attitude Scale.—The reliability of the attitude scale, measured by the split-half technique, and using the Spearman-Brown formula was +·871 (N=183); and, using the essay " Spare the rod and spoil the child " as the criterion, its validity was +·594 (N=57). Scores ranged from 0 to 11; the higher the score, the more unfavourable the attitude to corporal punishment.

Comparison of Colleges.—An analysis of variance showed that students from three different colleges differed in their attitudes to corporal punishment.

TABLE 1

Attitude Scores of Students from Three Different Colleges.

Group	N	M	S.D.	S.E.
Mixed College	86	5·89	1·326	0·143
Men's College	38	6·39	1·122	0·182
Women's College	54	6·99	0·763	0·104

Men and women students differed in their attitudes to corporal punishment even though they had been trained at the same college. The women were more unfavourable to corporal punishment than the men.

TABLE 2

ATTITUDE SCORES FOR MEN AND WOMEN STUDENTS.

Group	N	M	S.D.	S.E.	
All women	104	6·63	1·017	·0997	t=3·815
All men	74	5·91	1·378	·1603	1% level
Women—Mixed College	50	6·25	1·114	·1575	t=3·005
Men—Mixed College ..	36	5·39	1·433	·2388	1% level

There was no difference in the attitudes of either men or women who had been trained to teach different age groups of children.

Students in Training.—A test of significance of difference between means showed that there was no difference between the attitudes of those students who had attended mixed grammar schools and those who had attended girls' grammar schools. The attitude scores of these students bore no relation to their intelligence scores, or to their marks in the Practice or in the Principles of Education.

TABLE 3

CORRELATIONS OF ATTITUDE SCORES WITH OTHER MARKS.

N=57	Attitude Score
Intelligence N.I.I.P. Gp. Test 33	r=+·179
Practice of Education Mark	r=+·008
Principles of Education Mark	r=+·008

Three weeks first school practice did not affect the attitudes of either the infant or junior teachers. The attitudes towards corporal punishment of the schools where the students had practised had not affected attitude score changes.

Work at the children's play centre had not affected the students' attitudes, nor had the fourteen-week period of the ordinary college course.

All these results were found by testing changes in scores for significance of difference between means of related measures.

Young Teachers.—Fifty women students, whose attitudes had been measured on leaving college, were re-tested after 10 months of full-time teaching. The attitudes of the junior teachers had not changed, while the attitudes of the infant teachers had become significantly less unfavourable to corporal punishment. $(P < \cdot 01)$.

Further investigation showed that the significant changes were among those infant teachers who had taught classes of forty or more children (M= – ·78, S.D.= ·809, S.E.= ·1908), and those who had taught in residential areas (M= – ·86, S.D.= ·894, S.E.= ·2696).

Differences in the size of schools, and in the type of accommodation within them had not influenced the attitudes of either the infant or the junior teachers. These results were found by using tests of significance of difference between means of related measures.

Self-estimate of Attitude Score.—Students beginning their training were not clear about their views on corporal punishment. The correlation of their attitude scores with their own estimate of their position was + ·313. As the course proceeded, their estimate became more accurate—after the first school practice it was + ·498, and after the fourteen week experimental period, it was + ·443.

Students at the end of their training were much more aware of their own attitudes to corporal punishment (r= + ·746), and after 10 months of full-time teaching a correlation of the attitude scores of fifty young teachers with their own estimates yielded an r= + ·723.

The Thurstone Primary Mental Abilities and Academic Achievement in the Junior Forms of an English Grammar School.

By PHILIP MEREDITH.

(*Abstract of thesis submitted for the degree of Master in Education in the University of Manchester,* 1955.)

OBJECT.

The research examines the suitability of the Primary Mental Abilities Tests (P.M.A.) of L. L. and T. G. Thurstone as a guide to specialization in Science or Arts (languages) subjects in a grammar school.

METHOD.

The experimental population was composed of the whole of the 1950 and 1951 entries to a county grammar school—some 130 pupils (boys and girls) in each year. Mark lists in ten subjects (Table 1) were obtained from two first-year internal school examinations. These were standardized for each subject, so as to reduce the effects of different standards of marking by different teachers. The test booklet, *S.R.A. Primary Mental Abilities, Intermediate—Ages* 11-17*—Form AH*, was applied to the whole population, in form groups. Results of the Moray House tests of Intelligence, English, and Arithmetic, were available from the 11 plus examination.

Product-moment correlation coefficients were computed between all tests, viz., (i) the five Primary Mental Abilities, (ii) the three Moray House tests, and (iii) the ten school subject marks.

RESULTS.

Most of the correlation coefficients are low, but many are significant at the 1 per cent. level. These would have been higher if the test population had been less highly selected. Only the top 15 per cent. of the school population is selected for grammar school entry.

A.—*Relationships between Predictors.*

(i) *Intercorrelations between P.M.A. tests.*—These are lower than those obtained by American workers, and do not conflict strongly with the claim that the tests are measures of independent abilities.

(ii) *Intercorrelations between Moray House Tests.*—These are of the same order as those between the P.M.A. tests.

(iii) *Correlations between P.M.A. and Moray House Tests.*—On the whole, the values are of the order of size that would be expected on a priori grounds. S (Space) however, unexpectedly shows the highest correlation with Moray House I.Q., in contrast with the results of most American investigators who found it to give the lowest relationship with American tests of " general intelligence."

B.—*Relationships between the Criteria.*

Many of the intercorrelations between subject marks are of the same order as those given above. A few are considerably higher, being between subjects usually grouped together in the school curriculum, e.g., the sciences, and the mathematical subjects.

C.—*Relationships between Predictors and Criteria.*

(i) *Correlations between P.M.A. tests and subject marks.*

Verbal-meaning (V) correlates significantly with all subjects except Arithmetic and Algebra plus Geometry. The highest correlations were with English Language and English Literature.

Space (S) does not correlate significantly with any subjects, but is related to the mathematical and scientific subjects as distinct from arts subjects.

Reasoning (R) gives the highest correlations with Algebra-Geometry and Arithmetic, but with other subjects it gives small correlations—lower than might have been expected.

Number (N) correlates most highly with Arithmetic, and next highest, unexpectedly, with French. Smaller but significant positive correlations are obtained with Algebra-Geometry, English Language and English Literature. An unexpected, significant negative correlation is shown with Physics.

Word-fluency (W) correlates significantly with all the arts, but with Biology only among the science and mathematical subjects.

From these correlations and from the multiple correlations shown in Table 1, it appears that the P.M.A. tests give nearly as good prediction of general academic achievement as do the Moray House tests, but that they do not differentiate clearly between ability in science and in arts subjects, perhaps because in the lower school curriculum the sciences are descriptive and are, therefore, associated with ability in English. The P.M.A. tests discriminate clearly between the mathematical and the arts subjects; S, R and N indicate ability in Mathematics, and V and W predict achievement in the Arts.

(ii) *Correlations between Moray House Tests and subject marks.*—The Moray House Intelligence test correlates significantly with every subject except French. Moray House English correlates highest with English Language—much more so than do V and W—and with all other subjects except Geography, Arithmetic and Algebra-Geometry. Moray House Arithmetic correlates more highly with Arithmetic than does N and also correlates significantly with Algebra-Geometry and French, but with no other subject.

TABLE 1

MULTIPLE CORRELATION COEFFICIENTS OF COMBINATIONS OF THE P.M.A. AND OF THE MORAY HOUSE TESTS WITH SUBJECT MARKS.

	P.M.A. Tests VSRNW	Moray House Tests IEA
English Language	·460	·555
English Literature	·353	·284
Geography	·227	·237
History	·274	·263
Arithmetic	·322	·540
Algebra-Geometry..........	·307	·368
Chemistry	·207	·365
Physics	·354	·405
Biology	·363	·314
French	·343	·287

The multiple correlations shown in Table 1 indicate that the Moray House tests might well be preferable to the P.M.A. tests as an instrument of academic guidance. With either set, discrimination between mathematical and arts subjects is clearer than between science and arts subjects.

D.—*Factor Analysis of the* 18 × 18 *matrix of all the correlations.*

Four factors are indicated by the analysis: (i) a "mathematical-numerical" factor; (ii) an "English-verbal" factor; (iii) a "scientific-spatial" factor, and a fourth factor which discriminates between the school subjects and the psychological tests.

Television Viewing in a Secondary Modern School.

By W. W. MITCHELL.

(Abstract of a dissertation presented in part fulfilment of the requirements of the Diploma of Education in the University of Manchester, June, 1956.)

The investigation was carried out in 1955 into the amount of time spent watching television by 208 boys in the six first-year forms of a secondary modern school in Lancashire, and attempted to discover any differences between the viewers and the non-viewers.

A first questionnaire asked boys to give estimates of the number of times a week during the previous three months they had watched the children's, or the evening, television programme wholly or partly. Next, from copies of the *Radio Times* in which individual programmes were specially numbered, the boys recorded the programmes seen during the previous week. Where necessary, the replies include a " ½ " or " ¼ " to indicate the estimated proportion actually seen. Replies were anonymous but capable of identification, and from them the viewing time was calculated for each boy.

The eighteen boys who viewed most of all were matched in pairs on I.Q. and home background with eighteen boys who either did not view at all or who viewed very little. A comparison was made between these two groups on the results of a questionnaire given to the whole population on interests and habits.

Subjects were unaware that the questionnaire had any connection with television viewing, and replies were anonymous. This questionnaire asked for information on the number and type of books, newspapers, comics and periodicals (juvenile and adult) read regularly ; the number and type of hobbies currently followed ; membership of clubs and the number of visits made to them ; and the average number of visits made to the cinema each week. The replies of the two special groups were extracted after identification. Later from information obtained by interview and from school records the groups were compared as wholes and in pairs for chronological age, reading age, form positions, teachers' ratings on behaviour and attitude to work, attendance, participation in sport outside school, size of family, favourite school subjects and appearance in group photographs. The subjects thought that they had been chosen at random and were unaware of the existence of the two groups.

76 per cent. of all boys tested in the first questionnaire said that they had viewed television regularly during the previous three months, the minimum frequency included being once a fortnight. Although the upper forms had the highest proportion of viewers, those who viewed in the lower forms did so more often. During the particular week examined in the second questionnaire, the greatest amount of time spent viewing by any single boy was thirty hours. 47 per cent. of all boys viewed for an average of one hour each evening and 27 per cent. for and average of two hours. 50 per cent. of all boys saw the children's programme five times a week and 35 per cent. saw it six or seven times.

No significant differences were found between the two matched groups of 18 viewers and 18 non-viewers, except that, as might be expected, the viewers visited the cinema rather less often. Both groups read from one to twelve books during the term ; most of them read the *Express* and the *Mirror*, in which they liked the sport, crime and comic strip sections ; both groups found time to join an organisation such as the Scouts or an athletic club and to pursue a hobby (usually collecting or making things) and both were spread evenly over the various levels of ability and achievement. Their teachers were unable to tell the viewers from the non-viewers by their behaviour or their industry.

The Effects of Environment on the Reading Ability of Eight-year-old Children.

By SYBIL M. RIDEN.

(*Abstract of thesis submitted for the degree of Master in Education in the University of Manchester,* 1955.)

AIM OF THE INVESTIGATION.

The aim of this research was to study the effects of the home and school environments on children's ability to learn to read. The questions asked were :

(1) How do the standards of reading compare, at the age of eight years, in schools where children are living in (*a*) a good residential area, (*b*) a poor residential area, and (*c*) a village ?

(2) What are the main influences in these different environments that may account for any differences of standards found ?

" Reading ability " was considered to be the number of words that the child could read from Vernon's standardisation of Burt's Word Reading Test.

THE SUBJECTS.

Three schools were used, situated in areas that were " good residential ", " poor residential " and " village." Every child in each of these schools who was between the ages of seven-and-a-half and eight-and-a-half years in June, 1953, was included in the sample. The numbers in the three schools were :

Good residential area	48
Poor residential area	47
Village area	31

METHODS.

(1) *Testing.*—Every child was tested with :

(*a*) The Burt (re-arranged) Word Reading Test.
(*b*) The Otis Group Intelligence Scale, Primary Examination Form A (non-verbal).
(*c*) The Terman-Merrill adaptation of the Binet Intelligence Scale.

(2) *The Home Environmental Assessments.*—Information concerning the home conditions of the children was collected by means of questionnaires filled in by their teachers. This information was used to make scales whereby the reading ability of each child could be compared with certain aspects of his home conditions. Two scales were made ; a " Culture " scale, and an " Occupations " scale. The " Occupations " scale was based on the occupational level of the father.

(3) *The School Environment Assessments.*—Information concerning school environments was obtained in the following ways :

(*a*) Accounts given by the head teachers concerning general methods of teaching reading over the past three years.
(*b*) Accounts given by individual class teachers of detailed methods of teaching reading in their classes.
(*c*) Observations made by the investigator of equipment, books, apparatus, type of book corner and actual teaching method.
(*d*) Information gathered from time tables, programmes and syllabuses in the schools.

An assessment of teachers' attitudes to education was made by the investigator.

This information was used for comparing the three school environments by means of three-point scales weighted to assess aspects such as suitability of school buildings and teaching methods.

RESULTS.

The following mean scores were obtained :

	Good Residential Area	Poor Residential Area	Village
Reading Quotients (based on chron. age)	100·20	80·40	97·73
Intelligence Quotients (Terman-Merrill Scale)	110·94	95·35	106·76
Reading Quotients (based on mental age)	90·37	90·45	90·60
" Culture " (12-point scale)	9·06	5·04	8·74
" Occupations " (5-point scale with ' 5 ' rated as lowest point)	2·73	4·25	3·0
Attendance (percentage)	90·90	85·68	90·77

The following correlations between Reading Ability and other factors were found :

	Good Residential Area	Poor Residential Area	Village
Intelligence (individual test)	·693	·615	·776
Intelligence (Non-Verbal)	·525	·328	·507
Occupational Level	·118	·177	·354
" Culture "	·050	·347	·396
Attendance	·244	·136	·109

Analytical methods of teaching reading were used in the good residential school, the phonetic method in both the others.

SUMMARY OF CONCLUSIONS.

1.—Standards of reading ability of eight-year-old children, judged by mean reading quotients, were found to be higher in the good residential area than in the poor one. Standards in the village were higher than in the poor residential area and only a little lower than in the good residential area.

2.—The main factor influencing reading ability was found to be intelligence.

3.—When reading standards were assessed by reading quotients based on mental instead of chronological ages, there was no significant difference between mean reading quotients in any of the three areas.

4.—Degrees of culture were similar in the homes in the good residential area and the village, but much lower in the poor residential area. Culture in the home was significantly correlated with reading in the poor residential area and the village where standards of culture in the school were low. There was practically no correlation between these two factors in the good residential area, however, where cultural standards in the school were high. It is possible that a minimum degree of culture is sufficient to influence reading ability, and this might be supplied in a good school environment even when cultural standards in the homes are low.

The results of this work indicate the need for further research into ways in which the school environment could be planned to serve the needs of different kinds of residential area. It would be useful to investigate further the suggestion that a raising of cultural levels in schools could counteract the deleterious effects on the development of reading ability of low standards of culture in the home.

BOOK REVIEWS

HAHN, M. E., and McLEAN, M.S. (1955). *Counselling Psychology.* McGraw-Hill (2nd edit.), pp. 302, 35s. 6d.

This is a major revision of a standard text-book and gives an up-to-date account of the training, functions and techniques of counselling psychologists. The counsellor has no counterpart in British psychology, but undertakes some of the functions of tutors, probation officers, social workers and psychiatrists in addition to many associated with clinical, industrial and educational psychologists. His field of employment includes schools, universities, industry, social agencies of various types, clinics and private practice. He is concerned with the guidance of " individuals faced with *vocational-educational-personal* problems " and takes independent responsibility for his work, but only attempts short-term treatment of problems within the normal range.

The method of treatment is the 'counselling interview.' The authors claim that this should be based on learning theory applied to the reduction of the client's anxieties but, at the same time, insist that it be eclectic and range from straightforward advice to non-directive psychotherapy. To the present reviewer, the section devoted to treatment seems the least satisfactory.

Actual counselling, however, is preceded by thorough evaluation, in terms of trait and factor theory, of the client's aptitudes, abilities, interests and personality characteristics and the major part of the book is devoted to a valuable didactic and comprehensive review of methods of individual assessment, including sociometry and semantics. Here the authors occasionally lapse, despite their own warning, into the assumption that reliability and validity are terms which apply only to standardised tests.

The fact that counselling is concerned mainly with teenagers and young adults, renders much of the book inapplicable to child guidance. The counsellor is not a member of an interdisciplinary team and the client's family receive no guidance. *Those concerned with vocational guidance will find it a worthwhile book.*

H. GWYNNE JONES.

PAUL D. LEEDY (1956). *Reading Improvement for Adults.* McGraw-Hill, pp. viii+456, 34s.

One of the most popular forms of ' Teach Yourself ' manuals nowadays is the *Teach Yourself to Read Well*—at least in the U.S.A. The full flood has not yet struck this country, but here is an advance wave. If and when the earnest reader has worked carefully through all the exercises given in this book, he should at least be adept in tackling this kind of exercise; there is no evidence that he will have learnt to do anything else. For it is doubtful if many people in their ordinary reading practise the technique advocated here of reading a passage straight through with sufficient care to answer detailed questions on it afterwards without re-reading; most of us would prefer to see the questions first and then look for the answers. Neither has it ever been proved that exercises to " widen the eye span " improve ordinary reading—or indeed that the " eye span " as distinct from the span of assimilation has anything to do with the reading of adults. Some useful tips may perhaps be gained from the chapters on skimming and on critical reading; but those on " readability " and on " meeting the reading demands of business " inevitably suggest the importance of a companion volume on *Writing Improvement for Adults.*

The psychologist will find little to interest him in this book; and the educationist will, it is to be hoped, treat it with caution. M. D. VERNON.

VERNON MALLINSON (1956). *None can be called Deformed.* William Heinemann, pp. x+214., 12s. 6d.

Realistically and objectively, this book presents and analyses the histories of thirty-six crippled adolescents (twenty-seven boys, nine girls) drawn from a wide range of social and economic backgrounds in five different European countries.

It is an unusual book, less by reason of the author's modest minimizing of its importance as an exploratory study than because, although his declared aim has been to consider the problems raised from the crippled child's point of view, he has managed (unlike so many who have written about handicapped children) to avoid the common pitfall of identifying himself with his subjects to the point of suppression of the proper exercise of a mature adult judgment. His attitude is sympathetic in the truest sense, but he is prepared to admit that treatment which, in specific instances, gave excellent results might, in others, have produced anything but a happy outcome.

He considers examples of crippling of severer and of slighter degree, including those of congenital as well as of later, or traumatic origin. The material is well arranged. Brief individual stories are first presented according to type of handicap. Reference numbers are given and, in each case, a letter denoting allotment to one of three categories of personal-social adjustment (good, fair, poor) which seemed to have been achieved when the more critical phases of adolescence were over. Thus, in the course of reading, easy reference can be made to the concise analysis of each history given in the first appendix. Within the chapters, at the end of each section devoted to a particular form of crippling, there follows a critical note evaluating the results of the interplay of personality and circumstance.

Of the conclusions which Mr. Mallinson draws, some admittedly will come as no surprise to those experienced in dealing with handicapped children, whether crippled or otherwise. Others, however, and in particular his views on the place of Special Schools in our educational system, cannot fail to provoke much thought concerning the sometimes apparently conflicting claims of, on the one hand, what is more immediately scholastically expedient for the crippled child (as distinct from those otherwise handicapped) and, on the other, the ultimate degree of personal satisfaction and adjustment likely to be achieved if, from the start, he is accepted as a normal member of the community.

There are three appendices concerning respectively experiments in Belgium and France and the provision for handicapped children's education in Holland. A short but useful bibliography is provided. M. I. DUNSDON.

N. O'CONNOR and J. TIZARD (1956). *The Social Problem of Mental Deficiency.* Pergamon Press, pp. vii+182, 30s.

The present book is essentially a continuous account of the researches of the two authors, published previously in various journals and now set in perspective to other published work.

The reports are detailed and instructive, and may help in encouraging further the unsentimental, constructive re-thinking of the problem of subnormality, which is taking place nowadays. After a short and concise discussion of the size of the problem and the community services available, most of the remainder of the space is devoted to experiments and researches in institutional settings. The carefully designed investigations deal with various aspects of occupational training and occupational adaptation, and their results destroy many of the traditional prejudices and beliefs still found in text-books on Mental Deficiency.

The authors use up-to-date statistical techniques and many types of tests, but state quite clearly that, in individual clinical work, these scientific methods are yet of little value. (" . . . ratings have been more successful predictors than more objective tests " (p. 85) " . . . the possibility of successfully using tests, or other measurements for individual guidance is small at present " (p. 163).)

In view of the limitations of predictive techniques, the authors rightly advocate using every opportunity of testing the mentally defective patient in real life situations rather than relying either on test-patterns and scatter analysis, or on the performance and behaviour inside the hospital.

They support their argument with the findings from their own investigations on occupational training and by reference to the many published papers on this subject, which suggest strongly how minor a part is played by cognitive short-comings in comparison with emotional and consequent social handicap. This suggests that psychotherapy and counselling should have a definite place in the treatment of the unstable defective. The effectiveness and feasibility of these procedures have been denied by many workers steeped in traditional mental deficiency practice—but the authors give a full account of a successful experiment in psychotherapy under controlled conditions.

It is unavoidable that the factual account, the emphasis on surveys and the results of statistical investigations must disregard to a certain extent the human being, the individual patient. The wide range of personality types, which make work in the field of Mental Deficiency so interesting, is neglected in work of this kind, and we are offered instead a composite picture of certain aspects of mental deficiency not usually found in the text-books. This is drawn with a certain reasonable optimism, the more impressive perhaps because it is based on exact and measured observations, and because the authors are so careful in their conclusions.

The book is doubly valuable for its clear and instructive exposition of operational research designs and for the detailed evidence, which supports the present-day demand to train the defective for community living and not for " colony life."

H. C. GUNZBURG.

J. PIAGET and B. INHELDER (1956). *The Child's Conception of Space*, translated by F. J. Langdon and J. L. Lunzer. London : Routledge and Kegan Paul, pp. xii+490, 42s.

This latest translated publication of work from Piaget's research centre embodies material and interpretations of particular interest to teachers of geometry. Together with *The Child's Conception of Number*, it is a ' must ' for those seriously concerned with the beginnings of mathematical thinking. It also reveals much more clearly than most of Piaget's earlier monographs his basic psychological assumption that the child actively *constructs* his concepts through the mediation of manipulation and perception. Apart from this general point, the discovery that children first construct topological space through the discovery of proximity, separation, spatial succession, enclosure and continuity, and that only after this stage are they able to pass on to construction of projective and Euclidean concepts, will interest mathematics teachers.

In the first section of the book the authors deal with the child's construction of topological space and describe experiments on haptic perception, drawing of simple geometrical figures, building up of spatial order using beads on rods, enclosure by using knots, and continuity by sub-dividing lines and figures.

The second part of the book is devoted to the child's construction of projective space, which differs from topological space in that it embraces the perception of many figures in relation to each other and ' a point of view,' whereas the second refers only to the intrinsic properties of particular figures. The construction of projective space entails conservation of lines, angles and distances. By a series of ingenious experiments on projection, shadows, linear perspective, geometrical sections and the development of surfaces, the authors demonstrate how these notions arise in the cognitions of the pupil.

The third section brings us to the Euclidean concepts of parallelism similarity as conserving angles, the notion of horizontal and vertical, and finally experiments on the child's capacity to reconstruct diagrammatical lay-outs. The last chapter provides an interesting summary and discussion of the intuition of space of particular relevance to the teaching of geometry. The authors tell us that the ' intuition ' of space is " not a ' reading ' or apprehension of the properties of objects, but from the very beginning, an action performed on them." It is precisely because he enriches physical reality instead of merely abstracting from it " a set of ready-made structures " that the child's action provides *operational* schemata that eventually function in " a purely abstract deductive fashion." These experiments and views support the direction of the progress in the teaching of mathematics, in particular geometry, during the last twenty-five years. This has shown itself in the recognition of unformalised intuition as the beginnings of learning geometry, followed by an appeal to concrete experience leading lastly, when the pupil is more mature, to a formalised and schematised geometrical discipline. E. A. PEEL.

K. W. Spence (1956). *Behaviour Theory and Conditioning.* New Haven: Yale University Press; London: Geoffrey Cumberledge, Oxford University Press, pp. ix+262.

Professor Spence has produced an interesting critical account of empirical and theoretical material from studies in the S–R theory of learning. He justifies the laboratory study of learning on the grounds that we need to systematise our information about the conditions under which learning takes place. The ' theoretical-experimental psychologist ' tries to formulate laws " that will have a wide range of application to particular instances of learned behaviour." After an enlightening account of selective learning and instrumental and classical conditioning, he reviews empirical evidence obtained from different forms of learning curves and, in the following chapter, discusses the problem of basing a theoretical learning model upon these empirical laws. Then he discusses reinforcement and motivation, and finally applies his theorectical model to more complex learning situations.

The book has much to offer the specialist in S–R theories and experimentation, particularly as the author is largely concerned with the balance between empirical results and theoretical models. It has less to offer the general reader in educational psychology, for even in the last chapter the complex learning discussed is mainly that seen in animals, although there is a section on the rôle of anxiety as a drive in the human learning of paired associates. One valuable consequence of reading the book is to make one more aware of some of the pitfalls of applying the S–R learning model, based on animal experiments, too readily to the complex situations of normal human learning. On the other hand, if these models are applied as ' theoretical constructs ' then research may be stimulated into classroom learning.

The book is to be recommended to those interested in explanations of learning, and particularly to advanced students of educational psychology who may wish to investigate the more complex learning achieved in the training and education of children.

E. A. Peel.

PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

J. M. Stephens (1956). *Educational Psychology.* Constable, pp. xv+717, 45s.

J. E. Floud, A. H. Halsey, and F. M. Martin (1956). *Social Class and Educational Opportunity.* Heinemann, pp. 151, 12s. 6d.

J. M. Tanner and Barbel Inhelder (Edit.) (1956). *Discussions on Child Development.* Tavistock, pp. 271, 28s.

G. E. R. Burroughs (1957). *A Study of the Vocabulary) of Young Children.* Univ. of Birmingham, Inst. of Ed., Monograph I. Oliver and Boyd, pp. 104, 6s.

W. E. Flood (1957). *The Problem of Vocabulary in the Popularisation of Science.* Educ. Monograph II, Univ. of B'ham Inst. of Ed. Oliver and Boyd., pp. 121, 6s.

Paul Lafitte (1957). *The Person in Psychology : Reality or Abstraction.* Routledge and Kegan Paul, pp. x+233, 25s.

Dorothea L. Meek (1957). *Soviet Youth.* Routledge and Kegan Paul, pp.x. + 251, 28s.

Frank S. Caprio (1957). *Variations in Sexual Behaviour.* Calder, pp. x+344, 30s.

Paul Halmos (1957). *Towards a Measure of Man.* Routledge and Kegan Paul, pp. viii+250, 28s.

F. J. Schonell, I. G. Middleton and B. A. Shaw, *et al* (1957). *A Study of the Oral Vocabulary of Adults.* Univ. of Queensland and Univ. of London, pp. 174, 20s.

Helga Eng (1957). *The Psychology of Child and Youth Drawing.* Routledge and Kegan Paul, pp. 205, 50s.

Jon Eisenson and Mardel Ogilvie (1957). *Speech Correction in the School.* Macmillan, New York, pp. vi+294, 30s.)

Agatha H. Bowley (1957). *The Natural Development of the Child—Fourth Edition.* Livingstone, pp. xvi+206, 10s. 6d.

Henry Clay Lindgren (1957). *Psychology of Personal and Social Adjustment.* American Book Company, pp. viii+481, $4.50.

D. W. Winnicot (1957). *The Child and the Outside World.* Tavistock Publications and Ruth Borchard, pp. viii+190, 16s.

D. W. Winnicot (1957). *The Child and the Family.* Tavistock Publications, pp.x+147, 12s. 6d.

D. W. Hamlyn (1957). *The Psychology of Perception.* Routledge and Kegan Paul, pp. vii+120, 12s. 6d.

Peter Geach (1957). *Mental Acts.* Routledge and Kegan Paul, pp.x+136, 12s. 6d.

Anthony, C. Li (1957). *The History of Privately Controlled Higher Education in the Republic of China.* Catholic Univ. of America Press.

Margaret M. Clark (1957). *Left Handedness.* Scottish Council for Research in Education and London University Press, pp.xxii+212, 15s.

R. A. Hodgkin (1957). *Education and Change.* Oxford University Press, pp. 150, 9s. 6d.

J. Mainwaring (1957). *Psychology in the Classroom.* University of London Press, pp. x+118, 7s. 6d.

M. F. Cleugh (1957). *The Slow Learner.* Methuen, pp. 186, 13s. 6d.

A. T. Jersild (1957). *The Psychology of Adolescence.* Macmillan, pp. xii+438, 35s.

P. Chauchard (1956). *La Fatigue.* Paris : Presses Universitaires de France. pp. 128.

E. M. Caldwell (1956). *A Case of Spatial Inability in a Cerebral Palsied Child.* London : British Council for the Welfare of Spastics, pp. 28, 4s. 6d.

K. W. Spence (1957). *Behaviour Theory and Conditioning.* Yale and O.U.P., pp. 262, 36s.

J. E. W. Wallin (1956). *Mental Deficiency.* Vermont : Journal of Clinical Psychology, pp. xv+200, $5.00.

H. Roth (1955). *Autorität oder demokratisch erziehen ?* (Bedrohte Jugend-Drohende Jugend Series, No. 42.) Stuttgart : Ernst Klett Verlag, pp. 59, no price given.

W. Hemsing (1956). *Der Herr Sohn.* (Bedrohte Jugend-Drohende Jugend Series, No. 41.) Stuttgart : Ernst Klett Verlag, pp. 52, no price given.

K. Seelmann (1956). *Wie soll ich mein Kind aufklären ?* (Bedrohte Jugend-Drohende Jugend Series, No. 43.) Stuttgart : Ernst Klett Verlag, pp. 74, no price given.

W. Belser-Kuppers (1956). *Aber sie können sich nicht konzentrieren.* (Bedrohte Jugend-Drohende Jugend Series, No. 44.) Stuttgart : Ernst Klett Verlag, pp. 60, no price given.

F.Robaye (1956). *Niveaux D'Aspiration et D'Expectation.* Paris : Presses Universitaires de France, pp. viii+223, no price given.

M. Th. Maurette, H. Gratiot Alphandery, *et al* (1956). *Loisirs et Formation Culturelle de l'Enfant Rural.* Paris : Presses Universitaries de France, pp. xxiv+345, no price given.

M. Argyle (1957). *The Scientific Study of Social Behaviour.* London : Methuen and Co., Ltd., pp. 239, 21s.

A. Leon (1957). *Psychopédagogie de L'Orientation Professionnelle.* Paris : Presses Universitaires de France, pp. xi+131.

P. R. Farnsworth and others (1957). *Annual Review of Psychology.* Annual Reviews Inc., Calif., Vol. 8, pp. ix+502, $7.50.

E. V. Crane (1957). *My Minds and I.* The Christopher Publishing House, Boston, U.S.A., pp. 269, $4.00.

A. H. Bowley (1957). *The Young Handicapped Child.* Edinburgh and London E. and S. Livingstone, Ltd., pp. viii+127, 10s. 6d.

World Federation for Mental Health (1957). *Family Mental Health and the State.* Proceedings of the Eighth Annual Meeting, Istanbul, August, 1955. London : H. K. Lewis and Co., Ltd., pp. 165, 15 s.

La Psychanalyse (1956). *Mélanges Cliniques* (Publication No. 2). Paris : Presses Universitaires de France, pp. 323, no price given.

Educational Testing Service. *Annual Report,* 1955-56. Princeton, New Jersey, 1957., pp. 150.

Australian Council for Educational Research. *Twenty-sixth Annual Report,* 1955-56. Melbourne, Australia, pp. 60.

W. Litaner (1957). *Juvenile Delinquents in a Psychiatric Clinic.* London : Institute for the Study and Treatment of Delinquency, pp. 59.

County Councils Association, *et al* (1957). *School Television Broadcasting.* London : Association of Education Committees, pp. 11.

Brian Simon (Edit.) (1957). *Psychology in the Soviet Union.* Routledge and Kegan Paul, pp. viii+305, 32s.

Peter McKellar (1957). *Imagination and Thinking.* Cohen and West, pp. xi+219, 21s.

Michael Balint (1957). *Problems of Human Pleasure and Behaviour.* Hogarth Press and Institute of Psycho-analysis, pp. 300, 25s.

R. M. T. Kneebone (1957). *I work in a Secondary Modern School.* Routledge and Kegan Paul, pp. xvi+182, 12s. 6d.

D. J. O'Connor (1957). *The Philosophy of Education.* Routledge and Kegan and Paul, pp. vii+148, 10s. 6d.

Stanley Nisbet (1957). *Purpose in the Curriculum.* University of London Press, pp. 192, 10s. 6d.

Hans Selye (1957). *The Stress of Life.* Longmans, Green and Co., pp. xiv+324, 18s.

P. E. Vernon (edit.) (1957). *Secondary School Selection.* Methuen, pp. 216, 15s.

Ronald Fletcher (1957). *Instinct in Man.* London : Allen and Unwin, pp. 348, 40s.

Melanie Klein (1957). *Envy and Gratitude.* London : Tavistock Publications, pp. x+101, 12s. 6d.

John Rickman (1957). *Selected Contributions to Psycho-analysis.* London : Hogarth Press and the Institute of Psycho-analysis, pp. 411, 30s.

Chronicle of the World Health Organisation. Vol. II, Nos. 3, 4 and 5.

Calvin, S. Hall and Gardner Lindzey (1957). *Theories of Personality.* New York John Wiley and Sons, London, Chapman and Hall, pp. x+572, 52s.

Joan Saunders and Marjorie Napier (1957). *Spastics in Cheyne Walk.* London : Pitman Medical Publishing Co., pp. xiv+156, 20s.

Vernon W. Grant (1957). *The Psychology of Sexual Emotion.* Longmans Green and Co., pp. 270, 35s.

A. W. G. Ewing (edit.) (1957). *Educational Guidance of the Deaf Child.* Manchester University Press, pp. xvi+345, 45s.

Rupert Crawshay-Williams (1957). *Methods and Criteria of Reasoning.* London : Routledge and Kegan Paul, pp. 296, 32s.

Robert T. Oliver (1957). *The Psychology of Persuasive Speech* (second edit.). Longmans, Green and Co., pp. ix+466, 35s.

David Greenwood (1957). *Truth and Meaning.* New York : Philosophical Library, pp. viii+114, $3.75.

Group Work and Community Organisation. Papers presented at the 83rd Annual Forum of the National Conference of Social Work (1956). Columbia University Press ; London : Oxford University Press, pp. ix+108, 20s.

Lee J. Cronbach and Goldine C. Gleser (1957). *Psychological Tests and Personnel Decisions.* University of Illinois Press, pp. 165, $3.50.

Daniel H. Funkenstein and George H. Wilkie (1957). *Student Mental Health : An annotated Bibliography,* 1936-1955. World Federation for Mental Health, publ. H. K. Lewis and Co., Ltd., London, pp. vi+297, 20s.

C. A. Rogers (1956). *Measuring Intelligence in New Zealand.* Auckland University College, Monograph Series No. 2, pp. 127, no price given.

G. Spencer Brown (1957). *Probability and Scientific Inference.* Longman's Green and Co., Ltd., pp. viii+154, 15s.

ÉMILE CAILLE (1957). *Caractères et Écritures.* Presses Universitaires de France, pp. xi+290, no price given.

D. RUSSELL DAVIS (1957). *An Introduction to Psychopathology.* Oxford University Press, pp. vi+388, 30s.

A. HESNARD (1957). *Psychanalyse du Lien Interhumain.* Presses Universitaires de France, pp. 231, no price given.

J. O. ANCHEN (1956). *Frank Tate and His Work for Education.* Australian Council for Educational Research, pp. 222, no price given.

PAUL CHOUCHARD (1957). *Précis de Biologie Humaine : Les Bases Organiques du Comportement et de la Pensée.* Presses Universitaires de France, pp. xii+403, no price given.

INDEX TO VOLUME XXVII, 1957

Articles : PAGE

Attainment, Variability in : L. C. D. Kemp 211
Einstellung Rigidity, Intelligence and Teaching Methods : K. M. Miller .. 127
Essay Marking Experiments : Shorter and Longer Essays, by D. M. Edwards Penfold, Comments on : D. N. Lawley and A. E. G. Pilliner 142
Essay Marking Experiments, Reply to comments on : D. M. Edwards Penfold .. 145
Interests of Children (aged 10-11) in Books : J. D. Carsley 13
Marriage, the Prospect of : A study of the Attitudes towards Further Education of a Sample Group of Secondary Technical and Secondary Grammar School Leavers : M. Gallagher 24
Personality Traits in Seven-Year-Old Children, the Distribution of : A combined Psychological, Psychiatric and Somatotype Study : M. A. Davidson, R. G. McInnes, and R. W. Parnell 48
Piaget's Work, Recent Developments in : D. E. Berlyne 1
Projective Technique, The Objective Use of, Illustrated by a study of the Differences in Attitudes between Pupils of Grammar Schools and of Secondary Modern Schools : E. Bene 89
Reading Tests, A Note on the use of the Schonell and Gates in the First Year of the Junior School : M. L. Kellmer Pringle and M. D. Neale 135
School Journey, The Relation between and Social and Emotional Adjustment in Rural Infant Children : T. Lee 101
Schools for Maladjusted and Ordinary Boarding Schools, Differences between : M. L. Kellmer Pringle 29
Student Teachers, An Analysis of the Views of Supervisors on the Attributes of Successful Graduate : J. D. C. Robertson 115
Symposium : The Use of Essays in Selection at 11+ : IV.—Experimental Inquiries into the Use of Essay-type English Papers : D. A. Pidgeon and A. Yates .. 37
Symposium : The Development of Children's Moral Values : I.—Some Aspects of Moral Development in a Changing Society : J. Hemming 77
Symposium : Contributions to Intelligence Testing and the Theory of Intelligence :
    I.—On Defining Intelligence : T. R. Miles 153
    II.—Nature-Nurture Balance Sheets : H. Maddox 166
    III.—A Study of Cognitive Error through an Analysis of Intelligence Test Errors : P. J. Lawrence 176
    IV.—Intelligence and Age : Retesting with Twenty-Four Years' Interval : J. D. Nisbet 190
    V.—An Experiment on Practice, Coaching and Discussion of Errors in Mental Testing : A. W. Heim and K. P. Watts 199
Temperamental Characteristics Related to Disparity of Attainment in Reading and Arithmetic : R. Lynn 62

Summaries of Researches Reported in Degree Theses 68

Authors :

Bene, E. : The Objective Use of a Projective Technique, Illustrated by a Study of the Differences in Attitudes between Pupils of Grammar Schools and of Secondary Modern Schools 89
Berlyne, D. E. : Recent Developments in Piaget's Work 1
Carsley, J. D. : The Interests of Children (aged 10-11) in Books 13
Davidson, M. A., McInnes, R. G., and Parnell, R. W. : The Distribution of Personality Traits in Seven-Year-Old Children : A Combined Psychological Psychiatric and Somatotype Study 48
Edwards Penfold, D.M. : Reply to Comments on " Essay Marking Experiments " 145
Gallagher, M. : The Prospect of Marriage : A Study of the Attitudes towards Further Education of a Sample Group of Secondary Technical and Secondary Grammar School Leavers 24
Hemming, J. : Symposium : The Development of Children's Moral Values. I.—Some Aspects of Moral Development in a Changing Society 77
Heim, A. W., and Watts, K. P. : Symposium : Contributions to Intelligence Testing and the Theory of Intelligence. V.—An Experiment on Practice, Coaching and Discussion of Errors in Mental Testing 199
Kemp, L. C. D. : Variability in Attainment 211
Lawley, D. N., and Pilliner, A. E. G. : Comments on "Essay Marking Experiments : Shorter and Longer Essays " by D. M. Edwards Penfold 142

PAGE

Lawrence, P. J.: Symposium: Contributions to Intelligence Testing and the Theory of Intelligence. III.—A Study of Cognitive Error through an Analysis of Intelligence Test Errors. 176

Lee, T.: On the Relation between the School Journey and Social and Emotional Adjustment in Rural Infant Children 101

Lynn, R.: Temperamental Characteristics Related to Disparity of Attainment in Reading and Arithmetic 62

Maddox, H.: Symposium: Contributions to Intelligence Testing and the Theory of Intelligence. II.—Nature-Nurture Balance Sheets 166

Miles, T. R.: Symposium: Contributions to Intelligence Testing and the Theory of Intelligence. I.—On Defining Intelligence 153

Miller, K. M.: Einstellung Rigidity, Intelligence and Teaching Methods 127

Nisbet, J. D.: Symposium: Contributions to Intelligence Testing and the Theory of Intelligence. IV.—Intelligence and Age: Retesting with Twenty-four Years' Interval 190

Pidgeon, D. A., and Yates, A.: Symposium: The Use of Essays in Selection at 11+. IV.—Experimental Inquiries into the Use of Essay-Type English Papers 37

Pringle, M. L. Kellmer: Differences between Schools for the Maladjusted and Ordinary Boarding Schools 29

Pringle, M. L. Kellmer and Neale, M. D.: A Note on the Use of the Schonell and Gates Reading Tests in the First Year of the Junior School 135

Robertson, J. D. C.: An Analysis of the Views of Supervisors on the Attributes of Successful Graduate Student Teachers 115

NOTES TO CONTRIBUTORS

1.—The Editor is glad to consider articles in the fields of educational and child psychology, whether written from the theoretical, clinical, experimental or psychometric standpoints. It is, however, desirable that statistical matter (formulae, tables, discussions of techniques, etc.) should be kept to a minimum, so that the *Journal* will be of interest to statistically untrained readers. Such concepts as the mean, percentiles, standard deviation, correlation, chi-squared, standard error, critical ratio, significance and reliability, and general or group factors, can be employed in the text. But more complex concepts and methods should be explained, or confined to footnotes or an appendix, or published elsewhere.

Articles which have been, or are to be, published elsewhere cannot be accepted.

2.—*Length.* The usual range of length is from 2,500 to 4,500 words, and only very exceptionally will articles of more than 6,000 words be published. Short research reports of 500—1,000 words, such as outlines of theses, are welcomed. These should normally be submitted through the Head of the Department in which the research was carried out.

3.—*Arrangement.* Articles should be *preceded* by a summary or abstract of between 100 and 300 words, embodying the main conclusions. This should be typed directly after the title and author's name and place of work. A contents analysis, or list of headings, need *not* be included, but successive sections of the article should receive numbered headings, e.g. :

I. INTRODUCTION.
II. OUTLINE OF PREVIOUS RESEARCH.
III. PLAN OF THE INVESTIGATION.
    (i) *Subjects and materials.*
    (ii) *Experimental Procedure.*
IV. STATEMENT OF RESULTS.
V. DISCUSSION AND CONCLUSIONS.
VI. REFERENCES.

4.—*Typing.* All papers must be typed (double or 1½-spaced) on one side of quarto or foolscap paper. There should be margins of at least one inch all round. Authors should consult articles in current numbers of the *Journal*, and copy their conventions in respect of capitalising the title and main section headings, italicising author's place of work and sub-section headings, indentation of paragraphs, the form of references and tables, etc.

5.—*Footnotes* should be few in number, and should be indicated in the text by using the signs *, †, ‡, §.

6.—*References* should normally be indicated by date following the author's name in the text, e.g., Spearman (1904), or (Spearman, 1904, p. 160), or—where there are two or more publications in the same year—Burt (1939a). However, it is occasionally more convenient to number the references throughout, and to type numbers in the text above the line, e.g., Spearman.[4] When there are two or three references only, these may be placed in footnotes in the text, instead of in a separate section at the end.

The list of references at the end should always be in alphabetical order of author's names, and—for any one author—by order of date. The following standard form of references should be observed :

DEARBORN, W. F., and ROTHNEY, J. W. M. (1941). *Predicting the Child's Development.* Cambridge, Mass. : Sci.-Art.

(Note that Ltd., etc., after publisher's names, also the numbers of pages in the book, can be omitted.)
BURT, C. (1939a). The relations of educational abilities. *Brit. J. Educ. Psychol.*, 9, 55-71.

(Note the absence of inverted commas and of capital initials in the titles of articles. The number of the volume is in arabic, not roman numerals. The part of the journal is not given, but page numbers are included in full).

The World List of abbreviations (as given in the December issues of *Psychological Abstracts*) is used for journals.

7.—*Tables and figures* are costly to print, and should be used only when they save space and/or are essential for effective presentation. The same material should not be presented in a table *and* a graph. Each table or figure should be numbered (in arabic figures), on a separate sheet, and its approximate position indicated in the typescript. Drawings should be in Indian ink on heavy unruled paper, the lettering on graphs being in pencil. Lines in tables should be ruled in pencil. Each table or graph should have a short title, sufficient for understanding without searching in the text.

8.—*Corrections.* The Editor cannot consider for publication articles which are seriously deficient in presentation. The typescript submitted (after preliminary consultation, if desired) should represent the final form in which the author wishes the paper to appear. Since changes in proof are costly, out of all proportion to the original setting, authors must be charged for any changes, insertions or deletions (other than printer's errors).

9.—*Reprints.* Authors receive twenty-five copies of their papers free. Extra copies, in multiples of twenty-five, may be had at cost price if the order is given when the proof is returned.

NATIONAL INSTITUTE OF EDUCATION